Janssen Research & Development

Statistical Analysis Plan

Double-Blind, Randomized, Parallel-Group Study with Quetiapine Extended Release as Comparator to Evaluate the Efficacy and Safety of Seltorexant 20 mg as Adjunctive Therapy to Antidepressants in Adult and Elderly Patients with Major Depressive Disorder with Insomnia Symptoms Who Have Responded Inadequately to Antidepressant Therapy

Protocol 42847922MDD3005; Amendment 3; Phase 3

JNJ-42847922 (seltorexant)

Status:	Approved
Date:	17 Nov 2023
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-134321

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

Confidentiality Statement

The information provided herein contains Company trade secrets, commercial or financial information that the Company customarily holds close and treats as confidential. The information is being provided under the assurance that the recipient will maintain the confidentiality of the information under applicable statutes, regulations, rules, protective orders or otherwise.

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

VERSION HISTORY **5**

1. INTRODUCTION **9**
1.1. Objectives and Endpoints 9
1.2. Study Design 12

2. STATISTICAL HYPOTHESES **12**

3. SAMPLE SIZE DETERMINATION **13**

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS **13**

5. STATISTICAL ANALYSES **14**
5.1. General Considerations 14
5.1.1. Visit Windows 14
5.1.2. Analysis Phases 18
5.1.3. Imputation of Efficacy 18
5.1.4. Pooling Algorithm 19
5.2. Participant Dispositions 19
5.3. Primary Endpoint(s) Analysis 20
5.3.1. Definition 20
5.3.2. Estimand 1 21
5.3.2.1. Analysis Methods 22
5.3.2.1.1. Primary Analysis 22
5.3.2.1.2. Sensitivity Analysis 22
5.3.2.1.3. Supplementary Analysis 23
5.3.3. Estimand 1a 23
5.3.3.1. Analysis Methods 23
5.3.3.1.1. Primary Analysis 23
5.3.3.1.2. Sensitivity Analysis 23
5.3.3.1.3. Supplementary Analysis 23
5.4. Secondary Endpoint(s) Analysis 23
5.4.1. Key Secondary Endpoint 23
5.4.1.1. Definition 23
5.4.1.2. Estimand 1 24
5.4.1.3. Analysis Methods 25
5.4.1.3.1. Sensitivity Analysis 25
5.4.1.4. Estimand 2 26
5.4.1.5. Analysis Methods 27
5.4.2. Other Secondary Endpoint(s) 28
5.4.2.1. Time to Treatment Related Discontinuation of Study Drug 28
5.4.2.1.1. Definition 28
5.4.2.1.2. Analysis Methods 28
5.4.2.2. MADRS 29
5.4.2.2.1. Definition 29
5.4.2.2.2. Analysis Methods 29
5.4.2.3. MADRS-6 29
5.4.2.3.1. Definition 29
5.4.2.3.2. Analysis Methods 29
5.4.2.4. MADRS – WOSI (Without Sleep Item) 30
5.4.2.4.1. Definition 30
5.4.2.4.2. Analysis Methods 30
5.4.2.5. Patient Health Questionnaire - 9 Item (PHQ-9) 30

5.4.2.5.1. Definition 30
5.4.2.5.2. Analysis Methods 30
5.4.2.6. Remission Based on MADRS Total Score 31
5.4.2.6.1. Definition 31
5.4.2.6.2. Analysis Methods 31
5.4.2.7. Response Based on MADRS Total Score with Mild or Better Symptoms 31
5.4.2.7.1. Definition 31
5.4.2.7.2. Analysis Methods 31
5.4.2.8. Clinically Relevant Weight Gain 32
5.4.2.8.1. Definition 32
5.4.2.8.2. Analysis Methods 32
5.5. Tertiary/Exploratory Endpoint(s) Analysis 32
5.5.1. Patient Reported Outcome Measurement Information System-Sleep Disturbance (PROMIS-SD) Short Form 32
5.5.1.1. Definition 32
5.5.1.2. Analysis Methods 33
5.5.2. Patient Reported Outcome Measurement Information System-Fatigue (PROMIS-Fatigue) Short Form 33
5.5.2.1. Definition 33
5.5.2.2. Analysis Methods 34
5.5.3. Insomnia Severity Index (ISI) Patient Version 34
5.5.3.1. Definition 34
5.5.3.2. Analysis Methods 34
5.5.4. Patient Global Impression of Severity (PGI-S) (Sleep Symptoms) 35
5.5.4.1. Definition 35
5.5.4.2. Analysis Methods 35
5.5.5. Consensus Sleep Diary 35
5.5.6. Quality of Life in Depression Scale (QLDS) 35
5.5.6.1. Definition 35
5.5.6.1.1. Analysis Methods 36
5.5.7. European Quality of Life (EuroQol) Group, 5 Dimension, 5-Level questionnaire (EQ-5D-5L) 36
5.5.7.1. Definition 36
5.5.7.2. Analysis Methods 36
5.5.8. Clinical Global Impression – Severity (CGI-S) (Depression symptoms) 37
5.5.8.1. Definition 37
5.5.8.2. Analysis Methods 37
5.5.9. Patient Global Impression of Change (PGI-C) (Depression) 37
5.5.9.1. Definition 37
5.5.9.2. Analysis Methods 37
5.5.10. Sheehan Disability Scale (SDS) 38
5.5.10.1. Definition 38
5.5.10.2. Analysis Methods 38
5.5.11. Ruminative Response Scale (RRS) 38
5.5.11.1. Definition 38
5.5.11.2. Analysis Methods 39
5.5.12. Response without Clinically Relevant Weight Gain 39
5.5.12.1. Definition 39
5.5.12.2. Analysis Methods 39
5.5.13. Correlation between MADRS total score and PROMIS-SD T-score 39
5.5.14. Subgroup Analyses 39
5.6. Safety Analyses 40
5.6.1. Extent of Exposure 40
5.6.1.1. Intervention Compliance 41
5.6.2. Adverse Events 41
5.6.3. Additional Safety Assessments 42
5.6.3.1. Clinical Laboratory Tests 42
5.6.3.1.1. Homeostatic Assessment (HOMA) Modeling 43

5.6.3.2. Vital Signs and Physical Examination Findings 44
5.6.3.3. Electrocardiogram 45
5.6.3.4. Other Safety Parameters 46
5.6.3.4.1. Columbia Suicide Severity Rating Scale (C-SSRS) 46
5.6.3.4.2. Arizona Sexual Experiences Scale (ASEX) 47
5.6.3.4.3. Cognitive Assessment (Revere.D) 48
5.6.3.4.4. Physician Withdrawal Checklist - 20 items (PWC-20) 49
5.6.3.4.5. Menstrual Cycle Tracking 49
5.7. Other Analyses 50
5.7.1. Pharmacokinetics 50
5.7.2. Biomarkers 50
5.7.3. Health Economics 50
5.7.3.1. Healthcare Resource Use Questionnaire (HRUQ) 50
5.7.4. Definition of Subgroups 51
5.8. Interim Analyses 52
5.8.1. Independent Data Monitoring Committee (IDMC) 52

6. SUPPORTING DOCUMENTATION 53
6.1. Appendix 1 List of Abbreviations 53
6.2. Appendix 2 Changes to Protocol-Planned Analyses 55
6.3. Appendix 3 Demographics and Baseline Characteristics 56
6.4. Appendix 4 Protocol Deviations 57
6.5. Appendix 5 Prior and Concomitant Medications 58
6.6. Appendix 6 Adverse Events of Special Interest 59
6.7. Appendix 7 Medications of Special Interest 60
6.8. Appendix 8 Conversion of Raw Score to T-Score for PROMIS-SD 61
6.9. Appendix 9 Conversion of Raw Score to T-Score for PROMIS-Fatigue 62
6.10. Appendix 10 Scoring of QLDS Items 63
6.11. Appendix 11 Criteria for Treatment-emergent Markedly Abnormal Laboratory Values 65
6.12. Appendix 12 Analyses for Assessing and Mitigating the Impact of COVID-19 on Study Outcome 67
6.13. Appendix 13 Analyses Methods for Revere.D Data 68

7. REFERENCES 71

VERSION HISTORY

Table 1 – SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	16 Dec 2020	Not Applicable	Initial release
2.0	4 Apr 2023	Section 4: Added analysis sets FAS1CON and FAS2CON to exclude Ukrainian participants who were ongoing at the time of the Ukraine-Russia geopolitical situation. Added Follow-up analysis set. Section 5.1: Updated visit names for follow-up visits. Minor edits to the analysis windows. Section 5.3.2: Changed the population used for estimand 1 to FAS1CON. Section 5.3.2.1.1: Changed the stratification factor of country to region in CMH analysis. Section 5.3.2.1.2: Added sensitivity analyses excluding MVP participants, and analysis including country as a stratification factor as described in the protocol. Section 5.3.2.1.3: Added supplementary analysis excluding both Ukraine and Russia participants who were ongoing at the time of the geopolitical situation. Section 5.3.3: Changed the population used for estimand 1 to FAS2CON. Section 5.3.3.1.1: Changed the stratification factors to match estimand 1 for CMH analysis. Added a sensitivity analysis including country and baseline MADRS level as stratification factors as described in the protocol. Section 5.4.1: Specified the intercurrent event “switch of add-on treatment and/or underlying antidepressant” separately to three distinct intercurrent events	Finetuned the analyses for the pertinent analysis populations and made the analysis plan consistent with the protocol objectives in addition to other minor changes. See Appendix 2 for rationale for changes to primary endpoint analysis.

SAP Version	Approval Date	Change	Rationale
		to improve clarity. Clarified the summary measure. Section 5.4.2.1.2: Removed a stratified log-rank test and added a listing of participants evaluated by the adjudication committee. Replaced country with region in the Cox proportional hazards model. Section 5.4.2.6, 5.4.2.7: Changed the population used for analysis to FAS1CON and FAS2CON. Section 5.5.3.2: Added a frequency distribution of individual ISI item scores. Section 5.5.4.2: Changed analysis method for PGI-S from ANCOVA on ranked data to MMRM on actual data. In addition, similar change applied to CGI-S in Section 5.5.8.2 and PGI-C in Section 5.5.9.2 Section 5.5.5: Moved description of Consensus Sleep Diary to a SAP addendum. Section 5.5.12: Changed the population used for analysis to FAS1CON and FAS2CON. Section 5.5.14: Added unstratified CMH analysis for subgroups. Section 5.6: Updated safety analyses to be based on planned intervention received. Section 5.6.1: Updated compliance calculation for study intervention. Section 5.6.3.1: Clarified definition of treatment emergent markedly abnormal value. Clarified that only participants with baseline values that don't meet criteria or if baseline value is missing will be eligible for analyses of hepatic toxicity. Added shift tables for Glucose,	

SAP Version	Approval Date	Change	Rationale
		Triglycerides, Total, HDL and LDL Cholesterol. Section 5.6.3.3: Add abnormal limits for RR interval. Section 5.6.3.4.1: Clarified scoring rule for C-SSRS. Section 5.6.3.4.3: Added RCI analysis described in Appendix 13 Section 5.7.1: Updated PK section to reflect a listing instead of summary statistics. Section 5.7.4: Clarified populations to use for subgroup analysis. Updated region categorization. Section 5.8: Updated to reflect decision not to conduct an IA. Appendix 2: Updated with summary of analysis changed. Appendix 3: Removed category for WOCBP vs WONCBP vs men. Appendix 5: Added an additional analysis for concomitant medication data. Appendix 6: Updated preferred term based on MedDRA 25.1 Appendix 11: Removed parameter 'Lactic acid dehydrogenase (LDH)' as it's not collected in the study. Updated units for Hemoglobin A1c. Appendix 12: Specified that if number of non-site visits is small then sensitivity analysis may not be performed. Appendix 13: Description of RCI analysis of Revere.D added.	

SAP Version	Approval Date	Change	Rationale
Version 3.0	9 Nov 2023	Section 5.1: Minor edits to the analysis windows. Section 5.2: Add time to discontinuation plot by reason for discontinuation. Section 5.3: Added CMH analysis at each timepoint for the primary endpoint and summarize reasons for non-response. Section 5.4.1.3: Specified variance-covariance matrix structure for MMRM; Adjusted increment value for tipping point sensitivity analysis for weight. Section 5.4.1.5: Added rules for handling multiple intercurrent events at the same time. Section 5.4.2.6: Added CMH analysis for remission rates over time. Section 5.7.1: Removed PK analysis	Additional analyses in response to FDA feedback in addition to other minor changes.
Version 4.0	17 Nov 2023	Section 5.6.3.1.1 added.	To quantify the relative impact of seltorexant vs quetiapine on metabolic health.

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for study 42847922MDD3005.

1.1. Objectives and Endpoints

The primary and secondary objectives of MDD3005 are listed below. Major depressive disorder with insomnia symptoms (MDDIS) is defined as MDD with IS as a) moderate to severe IS by a patient version ISI total score of ≥15 at the end of screening and b) a positive response for IS (MDD symptoms Item 4) on the Structured Clinical Interview for DSM-5 Axis I Disorders Clinical Trials Version (SCID-CT). In addition, the clinician version ISI total score ≥15 is also required since it was used in the Phase 2 program.

Objectives	Endpoints
Primary	
• To assess the efficacy of seltorexant compared with quetiapine XR as adjunctive therapy to an antidepressant drug in treatment response in participants with MDDIS who have had an inadequate response to current antidepressant therapy with an SSRI or SNRI	• Difference in proportions of responders (≥50% improvement in MADRS total score from baseline) at Week 26 for seltorexant vs quetiapine XR.
Key Secondary	
• To assess the change in weight with seltorexant compared with quetiapine XR as adjunctive therapy to an SSRI or SNRI	• Change from baseline to Week 26 in weight with seltorexant vs quetiapine XR.
Secondary	
To assess the efficacy of seltorexant compared with quetiapine XR as adjunctive therapy to an SSRI or SNRI on the following:	
• Delaying time to potentially treatment related discontinuations	• Difference in time to study drug discontinuation for potentially treatment related reasons*
• Symptoms of depression	• Change from baseline to Week 26 in MADRS total score.
• Symptoms of depression by subscales – Core symptoms of depression – Without Sleep Item	• – Change from baseline to Week 26 in MADRS-6 total score. – Change from baseline to Week 26 in the MADRS without sleep item (MADRS-WOSI) total score.
• Patient-reported symptoms of depression	• Change from baseline to Week 26 in patient health questionnaire, 9-item (PHQ-9) total score.

Objectives	Endpoints
• Remission of depressive symptoms	• Proportion of participants with MADRS total score ≤12 at Week 26.
• Response and mild or better symptoms of MDD	• Proportion of participants with a ≥50% improvement in MADRS total score and MADRS ≤18 at Week 26.
• Clinically relevant weight gain	• Difference in proportions of participants having weight increase ≥7% at Week 26.
Exploratory	
To assess the efficacy of seltorexant compared with quetiapine XR as adjunctive therapy to an SSRI/SNRI on the following:	
• Patient-reported outcomes of sleep disturbance	• Change from baseline over time in sleep disturbance using the PROMIS-SD Short Form (8a) T-score.
• Patient-reported outcomes of fatigue	• PROMIS-Fatigue
• Patient-reported insomnia symptoms	• Change from baseline over time in the ISI total score.
• Patient-reported assessment of sleep symptoms	• Change from baseline over time in sleep symptoms using the PGI-S.
• Patient-reported sleep diary	• Change from baseline over time in subjective sleep parameters as measured by the CSD.
• Patient-reported depression-related quality of life	• Change from baseline over time in the QLDS total score.
• Patient-reported health-related quality of life	• Change from baseline over time in health-related quality of life and health status, as assessed by the European Quality of Life, 5 Dimension, 5-Level (EQ-5D-5L) questionnaire.
• Clinical symptom severity of depression	• Change from baseline over time in the Clinical Global Impression-Severity (CGI-S) score.
• Patient-reported assessment of change in depressive symptoms	• Change from baseline over time in depressive symptoms using the Patient Global Impression of Change (PGI-C).
• Patient-reported global functioning (work/school, social, and family life)	• Change from baseline over time in the Sheehan Disability Scale (SDS) total score.

Objectives	Endpoints
• Patient-reported rumination symptoms	• Change from baseline over time in the RRS total score.
Safety	
• To assess the safety and tolerability of seltorexant compared with quetiapine XR as adjunctive therapy to an SSRI/SNRI in participants with MDDIS	• Adverse events (AEs) including AEs of special interest (AESI) • Treatment discontinuations due to TEAEs • Laboratory values, including lipids and glucose, and ECG • Vital signs and physical examination • Weight, body mass index (BMI), and waist circumference • Suicidality assessment using the C-SSRS • Withdrawal symptoms assessment using the Physician Withdrawal Checklist 20-items (PWC-20) • Cognition assessment (Revere.D) • Patient-reported sexual functioning using Arizona Sexual Experiences Scale (ASEX)

*Note: See definition of potentially treatment-related discontinuations in Section 5.4.2.1.1

Additional exploratory objectives:

- To assess the efficacy and safety of seltorexant compared with quetiapine XR as adjunctive therapy to an SSRI or SNRI in treatment response without clinically meaningful weight gain based on:
    - Difference in proportions of participants who met MADRS response criteria (≥50% improvement in MADRS total score from baseline) and did not have weight increase of ≥7% at Week 26 for seltorexant vs quetiapine XR.
- To assess health care utilization with Health Resource Utilization Questionnaire (HRUQ)
- To identify diagnostic biomarkers and to investigate changes in MDD-related biomarkers (eg, HPA axis, metabolic function, and biomarkers of immune system activation) in relation to clinical response on depression symptoms and IS upon adjunctive treatment with seltorexant.

To identify genetic (eg, CYP genes) and other factors that may influence the safety or tolerability of seltorexant in participants with MDDIS.

Details about the analysis of biomarker and pharmacokinetic data pertaining to the additional exploratory objectives will be described in their respective analysis plans separately.

1.2. Study Design

This is a multicenter, DB, randomized, parallel-group, and active-controlled, 26-week, study to assess the efficacy and safety of 20 mg seltorexant compared with flexibly dosed quetiapine XR (150 or 300 mg) as an adjunctive therapy in adult and elderly participants with major depressive disorder with insomnia symptoms (MDDIS) (18 to 74 years of age, inclusive), who have had an inadequate response to current antidepressant therapy with an SSRI/SNRI.

This study will enroll approximately 720 participants, to target approximately 678 participants in the full analysis set 1 (FAS1). The planned dose of seltorexant is 20 mg and quetiapine XR is 150 or 300 mg.

The study will consist of 3 phases: a screening phase (up to 30 days), a DB treatment phase (26 weeks), and a posttreatment follow-up phase (7-14 day after the end of DB treatment phase for all participants, and up to 196 days from baseline for participants who stop study treatment early).

Participants will continue to take their baseline SSRI/SNRI antidepressant (at the same dose, without change, and at approximately the same time as prior to entering the study) from screening through the first follow-up visit.

2. STATISTICAL HYPOTHESES

This study is designed to primarily show that seltorexant is superior to quetiapine XR as an adjunctive MDD treatment in achieving a response criterion (defined as ≥50% improvement in MADRS total score from baseline) at Week 26.

If p_T and p_C are the proportion of participants meeting response criteria at Week 26, for seltorexant and quetiapine XR respectively, then the primary hypothesis can be written as follows:

H_0: $p_T - p_C \leq 0$

H_1: $p_T - p_C > 0$

Superiority can be concluded if the two-sided p-value for the testing of the above hypothesis is less than 0.05.

In addition, this study has a key secondary hypothesis:

- If μ_T and μ_c are the mean change in weight at Week 26 for seltorexant and quetiapine XR respectively, then the key secondary hypothesis can be written as follows:

  H_0: $\mu_T - \mu_c \geq 0$

H_1: $\mu_T - \mu_c < 0$

The multiplicity adjustment between testing for primary and key secondary hypotheses is described in Section 5.4.1.5

3. SAMPLE SIZE DETERMINATION

The enrollment will need to target a total of approximately 678 participants in the FAS1. The randomization will be in a 1:1 ratio to seltorexant: quetiapine XR. The following assumptions were used in the sample size calculation, based on clinical interpretation of available data:

- 31% response rate in seltorexant group at end of the 26-week DB phase
- 20.2% response rate in quetiapine XR group at end of the 26-week DB phase
- Participants who discontinue the DB treatment early are considered as non-responders
- Two-sided type I error level (alpha) of 0.05 (or equivalently, one-sided alpha of 0.025)
- Power of 90% to reject a null hypothesis: response rate in the seltorexant group is less than or equal to response rate in the quetiapine XR group, using a testing method based on normal approximation to the binomial distribution.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Analysis Sets	**Description**
Enrolled	All participants who sign the ICF.
Randomized	The randomized analysis set includes all participants who were randomized in the study.
Full Analysis Set 1 (FAS1)	The full analysis set 1 (FAS1) includes all randomized participants who received at least 1 dose of study intervention and had baseline MADRS total score ≥24 per IWRS
Full Analysis Set 1 Without Ukraine Participants Involved in Conflict (FAS1CON)	The full analysis set 1 without Ukraine (FAS1CON) includes all randomized participants who received at least 1 dose of study intervention and had baseline MADRS total score ≥24 per IWRS but excludes the Ukraine participants who were ongoing in the double-blind phase at the time of the Ukraine-Russian war in 2022.
Full Analysis Set 2 (FAS2)	The full analysis set 2 (FAS2) includes all randomized participants who received at least 1 dose of study intervention
Full Analysis Set 2 Without Ukraine Participants Involved in Conflict (FAS2CON)	The full analysis set 2 without Ukraine (FAS2CON) includes all randomized participants who received at least 1 dose of study intervention but excludes the Ukraine participants who were ongoing in the double-blind phase at the time of the Ukraine-Russian war in 2022.
Safety	The safety analysis set includes all randomized participants who received at least 1 dose of study intervention.
Pharmacokinetics Analysis Set	The PK analysis set is defined as participants who received at least 1 dose of active study intervention and have at least 1 valid blood sample drawn for PK analysis.
Follow-up Analysis Set	The follow-up analysis set includes all randomized participants who entered the follow-up phase.

5. STATISTICAL ANALYSES

5.1. General Considerations

5.1.1. Visit Windows

As participants do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to analysis visits. Listed below are the visit windows and the target days for each visit. The reference day is Study Day 1. If a participant has 2 or more actual visits in 1 visit window, the visit closest to the target day will be used as the protocol visit for that visit window. The other additional visit(s) will not be used in the summaries or analyses but they can be used for determination of clinically important endpoints. If 2 actual visits are equidistant from the target day within a visit window, the later visit is used.

All assignments will be made in chronological order. Once a visit date is assigned to a visit window, it will no longer be used for a later time point except for the endpoint. Listed below (Table 2) are the analysis visit windows and the target days for each visit defined in the protocol.

Table 2 – Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
MADRS	DB	3	Baseline	≤1	1
PROMIS-SD		6	Week 2	2 to 21	14
CGI-S		7	Week 4	22 to 35	28
PGI-S		8	Week 6	36 to 56	42
		9	Week 10	57 to 84	70
		10	Week 14	85 to 112	98
		11	Week 18	113 to 140	126
		12	Week 22	141 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
PHQ-9	DB	3	Baseline	≤1	1
		6	Week 2	2 to 28	14
		8	Week 6	29 to 70	42
		10	Week 14	71 to 126	98
		12	Week 22	127 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 70	42
			Week 14 EW	71 to 126	98

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
			Week 22 EW	127 to 168	154
			Week 26 EW	169 189	182
			Week 27-28 EW	190 to End of F/U	196
ISI (patient)	SCR	1	Screening 1	≤ -7	-19
		2	Screening 2	-6 to 1	-3
	DB	3	Baseline	≤1	1
		8	Week 6	2 to 70	42
		10	Week 14	71 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)		Week 27-28	1 to End of F/U	10
SDS QLDS EQ-5D-5L RRS HRUQ	DB	3	Baseline	≤1	1
		8	Week 6	2 to 70	42
		10	Week 14	71 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 27-28 EW	1 to End of F/U	196
PROMIS-Fatigue	DB	3	Baseline	≤1	1
		6	Week 2	2 to 28	14
		8	Week 6	29 to 70	42
		10	Week 14	71 to 126	98
		12	Week 22	127 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 27-28 EW	1 to End of F/U	196
PGI-C	DB	6	Week 2	2 to 21	14
		7	Week 4	22 to 35	28
		8	Week 6	36 to 56	42
		9	Week 10	57 to 84	70
		10	Week 14	85 to 112	98
		11	Week 18	113 to 140	126
		12	Week 22	141 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
CSD**	DB	3	Average Baseline	≤1	1
		8	Week 6	2 to 62	42
		13	Week 26	63 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
ECG	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		9	Week 10	2 to 104	70

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
		13	Week 26	105 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
Vital Signs	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		6	Week 2	2 to 21	14
		7	Week 4	22 to 35	28
		8	Week 6	36 to 56	42
		9	Week 10	57 to 84	70
		10	Week 14	85 to 112	98
		11	Week 18	113 to 140	126
		12	Week 22	141 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28 (F/U)	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
Weight	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		8	Week 6	2 to 56	42
		9	Week 10	57 to 84	70
		10	Week 14	85 to 112	98
		11	Week 18	113 to 140	126
		12	Week 22	141 to 168	154
		13	Week 26	169 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
Waist Circumference	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		8	Week 6	2 to 56	42
		10	Week 14	57 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
Menstrual Cycle Tracking***	SCR		Screening	<1	
	DB		Baseline	≤1	
			Week 1	1 to 7	
			Week 2	8 to 14	
			…	…	
			Week 26	176 to End of DB	
	F/U (DB)		Week 1 EW	1 to 7	
			Week 2 EW	8 to 14	
			…	…	
			Week 27-28 EW	183 to End of F/U	
Lab	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		8	Week 6	2 to 56	42
		10	Week 14	57 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
TSH/FT4, HbA1c	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		10	Week 14	2 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
C-SSRS	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		5	Day 8	2 to 11	8
		6	Week 2	12 to 21	14
		7	Week 4	22 to 35	28
		8	Week 6	36 to 56	42
		9	Week 10	57 to 84	70
		10	Week 14	85 to 112	98
		11	Week 18	113 to 140	126
		12	Week 22	141 to 168	154
		13	Week 26	169 to End of DB	182
	F/U (DB)	14	Follow-up 1	1	1
		15	Follow-up 2	2 to End of F/U	10
			Week 2 EW	2 to 28	14
			Week 6 EW	29 to 56	42
			Week 10 EW	57 to 84	70
			Week 14 EW	85 to 112	98
			Week 18 EW	113 to 140	126
			Week 22 EW	141 to 168	154
			Week 26 EW	169 to 189	182
			Week 27-28 EW	190 to End of F/U	196
ASEX	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		8	Week 6	2 to 56	42
		10	Week 14	57 to 140	98
		13	Week 26	141 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	
	F/U (DB)	15	Week 27-28	1 to End of F/U	10
PWC	DB	13	Endpoint (DB)	1 to End of DB	182
	F/U (DB)	14	Follow-up 1	1	1

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
		15	Follow-up 2	2 to End of F/U	10
Revere.D	SCR	1	Screening	<1	
	DB	3	Baseline	≤1	1
		9	Week 10	2 to 104	70
		13	Week 26	105 to End of DB	182
		DB final visit	Endpoint (DB)	2 to End of DB	

*Relative to the first day DB phase for analysis phases DB, F/U (DB). Assignment of visits to the FU (DB) will be made to the assessments that were recorded after double-blind treatment stop date. Participants who discontinue DB will have additional follow-up visits performed every 4 weeks until Day 196 - the follow-up data from these participants are assigned relative to the first day of DB phase with visit labels ending with EW ("EW" meaning early withdrawal).
** The analysis windows for CSD pertains to the average of the assessments that were performed 7 days prior to and including the study visit.
*** Analysis visits assignment for menstrual data are based on menstrual period start date (not actual visit date).

5.1.2. Analysis Phases

Double-blind Analysis Phase

The analysis reference start date of the double-blind analysis phase is the date of the first dose of double-blind study intervention. The analysis reference end date of the double-blind analysis phase is the maximum of the date of the last visit in the double-blind phase and date of completion or early withdrawal from the double-blind phase. For randomized participants who did not receive any study intervention in the double-blind phase, both analysis reference start and end dates are missing for the double-blind analysis phase.

Assignment of adverse events to double-blind analysis phase is described separately in Section 5.6.2.

Follow-up Phase

Start and end dates for the follow-up phase are only defined for participants who continued into the follow-up phase. The analysis reference start date of the follow-up analysis phase is the day after the reference end date for the double-blind analysis phase. The analysis reference end date of the follow-up analysis phase is the maximum of the last follow-up visit date or the disposition date at follow-up.

5.1.3. Imputation of Efficacy

Imputation method for missing data will include the following methods (Table 3).

Table 3: Imputation of Missing Efficacy Data

Imputation	Text
LOCF	Last observed value (excluding baseline value) will be used to replace the missing value regardless of study intervention exposure for treatment policy strategy for weight change.
Multiple Imputation (MI) method	Delta Adjustment

Imputation	Text
Non-Responder	Participants with missing values will be imputed as non-responders for analyses that use imputed data.
Non-Remitter	Participants with missing values will be imputed as non-remitters for analyses that use imputed data.

Imputation of total scores will be performed for the following efficacy scales as shown in Table 4 below. If the number of items with missing scores is greater than the maximum number of items presented in the table, the total score will be missing.

For the remaining efficacy assessments which require adding multiple item scores, the total score will be missing if any item score is missing.

Table 4: Imputation of Total Score for Efficacy Scales

Efficacy Scale	Total Number of Items	Maximum Number of Items That Can Be Missing *	Formula for Total Score
MADRS	10	1	Sum of non-missing item scores * (10 / number of non-missing items)
MADRS-WOSI	9	1	Sum of non-missing item scores * (9 / number of non-missing items)
PROMIS-SD	8	4	Sum of non-missing item scores * (8 / number of non-missing items). If the result is a fraction, the imputed total score will be rounded up to the nearest whole number.
PROMIS-Fatigue	8	4	Sum of non-missing item scores * (8 / number of non-missing items). If the result is a fraction, the imputed total score will be rounded up to the nearest whole number.
QLDS	34	6	Sum of non-missing item scores * (34 / number of non-missing items)

5.1.4. Pooling Algorithm

A total of 16 countries (Argentina, Belgium, Bulgaria, Canada, Czech Republic, Estonia, Latvia, Lithuania, Malaysia, Poland, Russia, Serbia, Slovakia, UK, Ukraine and USA) are to enroll participants. If low enrollment should occur in a country, it may be combined with other countries in a similar geographic region for analysis.

5.2. Participant Dispositions

The number of screened participants and reasons for screen failure will be summarized overall.

The number of participants in the following disposition categories will be summarized throughout the study by intervention group and overall for the all randomized, FAS1, FAS2, safety, and follow-up analysis sets:

- Participants randomized
- Participants who received study intervention

- Participants who completed, discontinued and reasons for discontinuation of study intervention
- Participants who completed, terminated and reasons for termination from the follow-up phase of the study
- Intercurrent events in the double-blind phase (for FAS1 and FAS2 only)

The distribution of the time to discontinuation of study intervention in double-blind phase will be displayed with Kaplan-Meier curves. A participant who discontinues study intervention at any time will be considered as an event, and the date of study intervention discontinuation will be used in the time to event calculation. A participant who completes the study intervention will be censored and the date of last dose of study intervention will serve as the time of censoring. In addition a Kaplan-Meier curve for time to discontinuation by reason for discontinuation will be provided for selected reasons.

Listings of participants will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who were unblinded during the study period
- Participants who were randomized yet did not receive study intervention.

Additional analyses of disposition data for assessing and mitigating the impact of COVID-19 on study outcome are presented in Appendix 12.

5.3. Primary Endpoint(s) Analysis

5.3.1. Definition

The primary efficacy endpoint is the proportion of participants who meet the response criteria (≥50% improvement in MADRS total score from baseline) at Week 26. Participants who discontinue treatment prior to Week 26 will not be considered meeting the response criteria.

The MADRS is a clinician-administered scale designed to measure depression severity and detects changes due to antidepressant intervention. (Montgomery 1979) The scale consists of 10 items, each of which is scored from 0 (item not present or normal) to 6 (severe or continuous presence of the symptoms). Higher scores represent a more severe condition. The MADRS evaluates reported sadness, apparent sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. The test exhibits high inter-rater reliability. The typical recall period for the MADRS is 7 days.

The MADRS total score is the sum of scores from individual question items at a given time point, and ranges from 0 to 60. Higher scores represent a more severe condition. Imputation of total score is presented in Section 5.1.3.

Negative changes in MADRS total score indicate improvement.

5.3.2. Estimand 1

Estimand 1 will be the primary estimand for submissions other than the EU dossier, and will be the supplementary estimand for the EU dossier. The analysis will be performed on FAS1CON.

Primary Trial Objective: To assess the efficacy of seltorexant compared with quetiapine XR as adjunctive therapy to an antidepressant in improving depressive symptoms in participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI or SNRI.

Estimand Scientific Question of Interest: What is the antidepressant benefit from seltorexant 20 mg versus quetiapine XR as adjunctive treatment to a SSRI or SNRI in adults and elderly participants with MDDIS, who have had an inadequate response to treatment with a SSRI/SNRI based on the proportion of participants who meet response criteria if the participants take study intervention as directed?

Estimand 1 is defined by the following 4 components:

Study Intervention:

- Experimental: Seltorexant dose of 20 mg as an adjunctive treatment to SSRI or SNRI
- Active Control: Flexibly dosed quetiapine XR as an adjunctive treatment to SSRI or SNRI

Population: Participants with MDDIS who have had an inadequate response to current antidepressant therapy with an SSRI/SNRI, as reflected by the inclusion/exclusion criteria and who have a baseline MADRS total score ≥ 24. For this estimand, the Ukraine participants who were ongoing on February 24, 2022 (approximately 13 of them) will be excluded since these participants did not have the opportunity to complete the double-blind phase (and hence did not have the opportunity to provide useful information for the primary endpoint) due to the geopolitical situation in Ukraine and Russia. If not excluded, these participants would automatically be counted as non-responders.

Variable: Proportion of participants who meet response criteria (≥50% improvement in MADRS total score from baseline) at Week 26. For calculating the proportion, the denominator is number of participants in the FAS within each intervention group. Note that participants who discontinue treatment prior to Week 26 will not be considered as meeting the response criteria.

Summary Measure: Difference in proportions between treatments.

Note that for the primary endpoint, participants who discontinue study drug early will not be considered as responders, by definition of this endpoint, its corresponding main estimand does not have the components for intercurrent events and strategies.

The population component requires a criterion of having baseline MADRS ≥24. The reason is that the study uses HDRS with remote rating for inclusion/exclusion during screening (as an effort to control potential rater inflation), however, at Day 1 (baseline), site rating of MADRS is needed as

the primary efficacy evaluation. In order to avoid the situation when HDRS rating has non-negligible discrepancy with MADRS rating, the population for the primary estimand is pre-planned to exclude participants with low MADRS (i.e., <24) at baseline since these participants do not represent the population of interest. This approach was discussed and agreed upon at the End of Phase 2 meeting with FDA (February 2020).

5.3.2.1. Analysis Methods

5.3.2.1.1. Primary Analysis

A Cochran-Mantel Haenszel (CMH) test stratified by geographic region, age group and baseline rumination level will be used to test the difference between seltorexant and quetiapine XR for the primary endpoint of response rate at Week 26.

A 2-sided 95% confidence interval for the intervention difference will be calculated based on the Wald statistic.

The point estimate and 2-sided 95% confidence interval will be provided for the relative risk of response.

The CMH analysis described above will also be performed for each time point.

The cumulative response rate, defined as the percentage of participants experiencing at least a given value of percent reduction from baseline to Week 26 in MADRS total score, will be presented graphically, for observed data. Additionally, response rates over time will be plotted.

The number and percentage of participants who achieve response will be summarized at each time point during the double-blind phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non-responders). The observed data will also be summarized to display the non-response reasons by MADRS non-response or treatment withdrawal.

Additionally, the response rate at Week 26 will be provided by study intervention and mode dose.

5.3.2.1.2. Sensitivity Analysis

A sensitivity analysis will be performed as in Section 5.3.2.1.1 but excluding participants with Major Protocol Violations (MPV), that is major protocol deviations that may impact efficacy such as violations of intended study population, errors in treatment assignment or use of excluded medication. Participants with a major deviation and/or a MPV will be determined prior to unblinding the data.

A sensitivity analysis will be performed as described in the protocol using the stratification level of country instead of region.

An unstratified CMH analysis will be performed as an additional sensitivity analysis.

5.3.2.1.3. Supplementary Analysis

A supplementary analysis will be performed as in Section 5.3.2.1.1 but excluding both Ukraine and Russian participants ongoing on Feb.24, 2022 due to the concern of potential impact of the geopolitical situation in Ukraine and Russia on the study outcome.

5.3.3. Estimand 1a

Estimand 1a will be the primary estimand for the EU dossier, and will be the supplementary estimand for submissions other than the EU dossier. The analysis will be performed on FAS2CON.

Estimand 1a is similar to Estimand 1 (Section 5.3.2) in all but the population component. The population under this estimand includes participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria (i.e., regardless of the baseline MADRS total score). For this estimand, the Ukraine participants who were ongoing on February 24, 2022 (approximately 13 of them) will be excluded since these participants did not have the opportunity to complete the double-blind phase (and hence did not have the opportunity to provide useful information for the primary endpoint) due to the geopolitical situation in Ukraine and Russia. If not excluded, these participants would automatically be counted as non-responders.

5.3.3.1. Analysis Methods

5.3.3.1.1. Primary Analysis

The same analysis for Estimand 1 will be applied to Estimand 1a.

5.3.3.1.2. Sensitivity Analysis

The same sensitivity analyses described for Estimand 1 will be performed for Estimand 1a.

5.3.3.1.3. Supplementary Analysis

A supplementary analysis will be performed as in Section 5.3.3.1.1 but excluding both Ukraine and Russian participants ongoing on Feb.24, 2022 due to the concern of potential impact of the geopolitical situation in Ukraine and Russia on the study outcome.

5.4. Secondary Endpoint(s) Analysis

5.4.1. Key Secondary Endpoint

The primary analysis of the key secondary endpoint will be based on full analysis set 1 (FAS1) using weight collected during the double-blind phase for submissions other than the EU dossier. For the EU dossier, the primary analysis will be based on full analysis set 2 (FAS2).

5.4.1.1. Definition

The key secondary endpoint is change in weight from baseline to Week 26.

The following estimands are defined for the key secondary endpoint: Estimand 1 using the FAS1 analysis set, and Estimand 2 using the FAS2 analysis set. For submissions other than the EU dossier, the primary estimand is Estimand 1, and the supplementary estimand is Estimand 2. For the EU dossier, the primary estimand is Estimand 2, and the supplementary estimand is Estimand 1.

The sections below describe the primary and sensitivity analyses performed for each primary and supplementary estimand.

5.4.1.2. Estimand 1

Primary Trial Objective: To assess the weight change of seltorexant compared with quetiapine XR as adjunctive therapy to an antidepressant in improving depressive symptoms in participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI or SNRI.

Estimand Scientific Question of Interest: What is the weight change benefit from seltorexant 20 mg versus quetiapine XR as adjunctive treatment to a SSRI or SNRI in adults and elderly participants with MDDIS, who have had an inadequate response to treatment with a SSRI/SNRI based on the change in weight if the participants take study intervention as directed?

Estimand 1 is defined by the following 5 components:

Study Intervention: same as Estimand 1 for the primary endpoint

Population: same as Estimand 1 for the primary endpoint

Variable: Change in weight from baseline to Week 26

Summary Measure: Difference in means of the variable between intervention groups.

Intercurrent events and their corresponding strategies:

Intercurrent Events	Name of Strategy for Addressing Intercurrent Events and Its Description
1. Treatment discontinuation of add-on study intervention only	Hypothetical strategy: as if the intercurrent event had not occurred
2. Treatment discontinuation of both underlying antidepressant and add-on study drug	Hypothetical strategy: see above
3. Switch of add-on study drug only	Hypothetical strategy: see above
4. Switch of underlying antidepressant only	Hypothetical strategy: see above
5. Switch of add-on study drug and underlying antidepressant	Hypothetical strategy: see above

5.4.1.3. Analysis Methods

Under Estimand 1 for weight change, the comparison between seltorexant and quetiapine in weight change will be performed using the appropriate contrasts in a MMRM, with main comparison at Week 26. The MMRM will include country, age group (adults [<65 years] and elderly [≥65 years]), baseline rumination level (RRS total score <54, ≥54), time, treatment and treatment by time interaction as factors, and baseline weight as a covariate. The within-subject covariance between visits will be estimated via an unstructured variance-covariance matrix. In case of convergence problems, alternative variance-covariance structures will be tried in the following order, with the first structure that converges being used in the analysis: heterogeneous Toeplitz, standard Toeplitz, and AR(1) with separate subject random effect. Comparisons between seltorexant and quetiapine XR at each timepoint including Week 26 will be performed using the appropriate contrast.

Additionally, descriptive statistics of the change from baseline to each postbaseline time point in the double-blind phase will be presented for weight change by intervention group and mode dose.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.4.1.3.1. Sensitivity Analysis

For the weight change, if the primary analysis is significant, delta adjustment with a tipping point will be conducted as a sensitivity analysis.

The delta adjustment multiple imputation method will be implemented on weight change if the results from the primary analysis show a significantly greater decrease in weight at Week 26 in seltorexant compared to quetiapine XR.

This method will employ the following 3 steps:

Step 1 – Multiple Imputation (MI)

If there are participants with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using methods such as the MCMC (Markov Chain Monte Carlo) method. 500 imputations will be performed to create 500 datasets which now have monotone missing (ie., missing data after the participant experienced an intercurrent event) data pattern.

Analysis assumptions: Missing at random (MAR) is assumed for intermediate missing data (i.e., missing data between non-missing observations). Monotone missing data will first be imputed by MAR-based MI regression, and the imputed scores in the experimental intervention group will be adjusted to be worse than the other participants in same group with non-missing data as discussed below.

The imputed values will be adjusted by adding δ_C to the imputed values for participants randomized to the control group and adding δ_A to the imputed values for participants randomized

to the experimental intervention group. Delta-adjusted fully imputed datasets will be generated for different combinations of δ_c and δ_A values as defined below:

- $\delta_c = 0$ and $\delta_A = 0$ *to* Δ^* in increments of 0.5 (experimental group-only adjustment analysis)
  - Adding positive values results in higher (worse) weight. Δ^* represents the adjustments leading to the 'tipping point', so the smallest delta adjustments values at which conclusions change from *favorable* (i.e. statistically significant: 2-sided p-value ≤0.05 in favor of seltorexant) to *unfavorable* (fail to reject the null hypothesis of no intervention difference).

- Imputed values in both the experimental and the control groups will be adjusted using a range of delta values, and delta-adjusted fully imputed datasets will be generated for each combination.

These methods will be applied to all data not used after the intercurrent events under hypothetical strategy. In addition, another version of these methods will be implemented, where the delta adjustments will be applied to all data not used after the intercurrent events under hypothetical strategy except for those caused by discontinuation reasons due to COVID-19, and for discontinuation reasons not related to study intervention including Lost to Follow-Up, Withdrawal by Subject or Other.

Step 2 – Analysis

Same MMRM analysis as described for the key secondary analysis (Section 5.4.1.3) will be performed for each set of the adjusted fully imputed datasets.

Step 3 – Pooling

Rubin's methodology will be applied to the MMRM results from the 500 imputed datasets to produce final inferences. (Rubin 1987)

Between-group comparisons to quetiapine XR at Week 26 (e.g. 2-sided p-values, point estimates for intervention difference) will be displayed graphically for each considered δ_A, up to the 'tipping point' adjustment. For analysis of the imputed datasets using a range of delta-adjustments on both experimental and control groups, a tipping point two-way map will be generated.

5.4.1.4. Estimand 2

All components as described under Estimand 1 (Section 5.4.1.2) apply to Estimand 2 except for the study population and strategy for addressing the intercurrent events:

Population:

Participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria (i.e., regardless of the baseline MADRS total score).

Intercurrent events and their corresponding strategies:

Intercurrent Events	Name of Strategy for Addressing Intercurrent Events and Its Description
1. Treatment discontinuation of add-on study intervention only	Treatment policy strategy: all observed values of the endpoint are used regardless of whether or not the participant had experienced this intercurrent event
2. Treatment discontinuation of both underlying antidepressant and add-on study drug	Hypothetical strategy: as if the intercurrent event had not occurred
3. Switch of add-on study drug	Hypothetical strategy: see above
4. Switch of underlying antidepressant	Hypothetical strategy: see above
5. Switch of add-on study drug and underlying antidepressant	Hypothetical strategy: see above

For participants experiencing multiple intercurrent events at the same time, intercurrent events in categories 3-5 will override the rest. For example, if a patient starts a new add-on study drug the day after taking the last dose of their previous add-on study drug, this intercurrent event will be considered a switch of add-on study drug, and not a discontinuation. However, if more than one day passes between the last dose of their previous add-on medication and the first dose of the new add-on medication, this situation will be considered a discontinuation of add-on study drug intercurrent event followed by a switch of add-on study drug intercurrent event.

5.4.1.5. Analysis Methods

Under Estimand 2, weight data collected after treatment discontinuation of add-on study drug only are planned to be included in the analyses, and the comparison between intervention groups will be performed using the appropriate contrast based on an analysis of covariance (ANCOVA) model using last observation carried forward (LOCF) data. The model will include factors for treatment, country, age group, baseline RRS group, baseline MADRS group, and baseline weight as a covariate.

A fixed sequence testing procedure will be applied to control the familywise error rate (FWER) at two-sided 0.05 level accounting for multiplicity due to the primary and key secondary efficacy endpoint. Per the testing sequence, the primary endpoint will be tested first, at the two-sided 0.05 level. If the hypothesis corresponding to the primary endpoint is rejected, then the key secondary endpoint will be tested at the same alpha level. If the hypothesis corresponding to primary endpoint is not rejected, then the testing will stop.

The multiplicity adjustment will be applied to the primary estimands for primary and key secondary endpoints within each dossier. For submissions other than EU dossier, the primary estimands for primary and key secondary endpoints will be Estimand 1 (using FAS1CON for primary endpoint, FAS1 for key secondary endpoint). For EU dossier, the primary estimand for primary endpoint will be Estimand 1a (using FAS2CON) and the primary estimand for key secondary endpoint will be Estimand 2 (using FAS2).

5.4.2. Other Secondary Endpoint(s)

Unless stated otherwise, the following analyses will be performed on the FAS1 and FAS2 populations. Unless specified, data collected after a participant discontinues study intervention will not be included in these analyses.

5.4.2.1. Time to Treatment Related Discontinuation of Study Drug

5.4.2.1.1. Definition

Time to discontinuation of study drug is defined as the number of days from the first dose of study drug to the last dose of study drug. Participants who complete double-blind treatment are not considered to have discontinued.

Potential treatment related discontinuations are assessed on an ongoing basis throughout the study treatment period. These are defined as all study drug discontinuations excluding the potentially non-treatment related discontinuations (eg, loss of insurance for antidepressant therapy, movement/travel out of the area, change of work-schedule being unable to accommodate visit schedule, family circumstances, site terminated by sponsor, study terminated by sponsor). Discontinuations due to AE or lack of efficacy or product quality complaints are considered potentially treatment related. For unclear or not well-defined reasons, an independent blinded adjudication review committee will be consulted.

5.4.2.1.2. Analysis Methods

The cumulative distribution function of time to potentially treatment related discontinuation of study drug will be estimated by the Kaplan-Meier method. Time to discontinuation will be summarized (median, 25th percentile, and 75th percentile and their 95% confidence intervals, if estimable, as well as the number of discontinuations, and the number of censored participants) by intervention group. Standard error estimates will be based on Greenwood's formula.

The estimate of the hazard ratio and its 2-sided 95% confidence interval will be provided based on a Cox proportional hazards model with intervention group, region, age group and baseline rumination level as factors. In addition, a similar model with intervention group as the only factor will be carried out.

The distribution of the time to discontinuation of study agent will be displayed with Kaplan-Meier curves. Participants who discontinued the double-blind phase at any time due to a potentially treatment related reason will be considered an 'Event' and their date of last dose of study agent based on the exposure dataset will be used in the time to event calculation. Participants who discontinue due to non-treatment related reasons, or complete the double-blind phase will be censored and the date of last dose of study agent based on the exposure dataset will serve as the time of censoring.

A listing of participants who discontinued and whether they were evaluated by the adjudication committee or not will be provided.

5.4.2.2. MADRS

5.4.2.2.1. Definition

A secondary endpoint is change from baseline to Week 26 in MADRS total score. For the full definition of MADRS, see Section 5.3.1.

5.4.2.2.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented for MADRS total score by intervention group.

MADRS total score will be analyzed by a Mixed-Effect Model for Repeated Measures (MMRM) based on observed case. The fixed terms included in the model will be intervention group, country, age group (adults [<65 years] and elderly [≥65 years]), baseline rumination level (RRS total score <54, ≥54), time, and time-by-intervention interaction, and baseline MADRS total score as a covariate. The within-subject covariance between visits will be estimated via an unstructured variance-covariance matrix. In case of convergence problems, alternative variance-covariance structures will be tried in the following order, with the first structure that converges being used in the analysis: heterogeneous Toeplitz, standard Toeplitz, and AR(1) with separate subject random effect. Comparison between seltorexant and quetiapine XR at all time points will be performed using the appropriate contrast. Difference in least square means and 2-sided 95% CI will be presented.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

Additionally, descriptive statistics of the change from baseline to each postbaseline time point in the double-blind phase will be presented for MADRS total score by intervention group and mode dose.

5.4.2.3. MADRS-6

5.4.2.3.1. Definition

The 6-item MADRS is a clinician-administered scale designed to measure the core symptoms of depression severity and detects changes due to antidepressant intervention. (Montgomery 1979) It is a subset of MADRS (10-item). The MADRS scale is a validated, reliable scale and acceptable to regulatory health authorities as a primary scale to determine efficacy in major depression.

The MADRS-6 subscale score is the sum of scores for the following MADRS items: reported sadness, apparent sadness, inner tension, lassitude, inability to feel, and pessimistic thoughts.

Negative changes in MADRS-6 total score indicate improvement.

5.4.2.3.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented for MADRS-6 total score by intervention group.

The change from baseline in MADRS-6 total score at all time points will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline MADRS-6 total score as the covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.4.2.4. MADRS – WOSI (Without Sleep Item)

5.4.2.4.1. Definition

MADRS-WOSI is defined as the full MADRS without the sleep item. (Montgomery 1979)

The MADRS-WOSI total score is the sum of scores of the remaining 9 items at a given time point, ranging from 0 to 54. Higher scores represent a more severe condition. Imputation of total score is presented in Section 5.1.3.

Negative changes in MADRS-WOSI total score indicate improvement.

5.4.2.4.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented for MADRS-WOSI total score by intervention group.

The change from baseline in MADRS-WOSI total score at all time points will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline MADRS-WOSI total score as the covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.4.2.5. Patient Health Questionnaire - 9 Item (PHQ-9)

5.4.2.5.1. Definition

The 9-item Patient Health Questionnaire - 9 Item (PHQ-9) scale scores each of the 9 symptom domains of the DSM-5 MDD criteria and it has been used both as a screening tool and a measure of response to intervention for depression. Each item is rated on a 4-point scale (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day). The participant's item responses are summed to provide a total score (range of 0 to 27), with higher scores indicating greater severity of depressive symptoms. The recall period is 2 weeks.

Negative changes in PHQ-9 total score indicate improvement.

5.4.2.5.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented for PHQ-9 total score by intervention group.

The change from baseline in PHQ-9 total score at Week 26 will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline PHQ-9 total score as the covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.4.2.6. Remission Based on MADRS Total Score

The following analysis will be performed on the FAS1CON and FAS2CON analysis sets.

5.4.2.6.1. Definition

A participant is defined as a remitter at a given time point if the MADRS total score is [≤12] at that time point. Participants who do not meet such criterion will be considered as non-remitters. Imputation of missing remission status is presented in Section 5.1.3.

5.4.2.6.2. Analysis Methods

The difference between intervention groups in remission rates will be analyzed at all time points using the same CMH model as described for the primary endpoint.

The number and percentage of participants who achieve remission will be summarized at each time point during the double-blind phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non-remitters).

Remission rates over time will be plotted.

5.4.2.7. Response Based on MADRS Total Score with Mild or Better Symptoms

The following analysis will be performed on the FAS1CON and FAS2CON analysis sets.

5.4.2.7.1. Definition

A secondary endpoint is the proportion of participants who meet the response criteria with mild or better symptoms (≥50% improvement in MADRS total score from baseline and MADRS total score ≤ 18) at Week 26.

5.4.2.7.2. Analysis Methods

The number and percentage of participants who achieve response with mild or better symptoms will be summarized at each time point during the double-blind phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non-responders).

5.4.2.8. Clinically Relevant Weight Gain

5.4.2.8.1. Definition

A participant is defined to have clinically relevant weight gain at Week 26 if the percent increase from baseline is ≥ 7%.

5.4.2.8.2. Analysis Methods

The number and percentage of participants who show clinically relevant weight gain will be summarized over time by intervention group and mode dose.

5.5. Tertiary/Exploratory Endpoint(s) Analysis

Exploratory analyses will be performed for the endpoints described in the sections below for FAS1 and FAS2 analysis sets.

5.5.1. Patient Reported Outcome Measurement Information System-Sleep Disturbance (PROMIS-SD) Short Form

5.5.1.1. Definition

Developed under a National Institutes of Health (NIH) initiative, the Patient Reported Outcomes Measurement Information System captures self-reported, qualitative health aspects in the domains of physical, mental, and social health. (Yu 2011)

The PROMIS-SD (Short Form 8a) subscale consists of a static 8-item questionnaire. The PROMIS-SD instruments assess self-reported perceptions of sleep quality, sleep depth, and restoration associated with sleep. This includes perceived difficulties and concerns with getting to sleep or staying asleep, as well as perceptions of the adequacy of and satisfaction with sleep. Sleep Disturbance does not focus on symptoms of specific sleep disorders, nor does it provide subjective estimates of sleep quantities (total amount of sleep, time to fall asleep, amount of wakefulness during sleep). The Sleep Disturbance short form is universal rather than disease-specific. It assesses sleep disturbance over the past 7 days.

Each question has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For the 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40. Lower scores indicate less sleep disturbance. The "direction" of the responses is not the same for all questions, i.e., sometimes a response of "not at all" indicates more sleep disturbance and sometimes a response of "not at all" indicates less sleep disturbance.

"My sleep quality was" ranges from 5=very poor to 1=very good
"My sleep was refreshing" ranges from 5=not at all to 1=very much
"I had a problem with my sleep" ranges from 1=not at all to 5=very much
"I had difficulty falling asleep" ranges from 1=not at all to 5=very much
"My sleep was restless" ranges from 1=not at all to 5=very much
"I tried hard to get to sleep" ranges from 1=not at all to 5=very much

"I worried about not being able to fall asleep" ranges from 1=not at all to 5=very much
"I was satisfied with my sleep" ranges from 5=not at all to 1=very much

A score can be approximated if a participant skips a question. Imputation of total score is presented in Section 5.1.3.

The total raw score or pro-rated raw score will be converted into a T-score for each participant based on the table in Appendix 8. The T-score rescales the raw score into a standardized score with a mean of 50 and an SD of 10.

5.5.1.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline timepoint in the double-blind phase will be presented for PROMIS-SD raw score and the T-score by intervention group.

The change from baseline in PROMIS-SD T-score over time will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline PROMIS-SD T-score total score as the covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.5.2. Patient Reported Outcome Measurement Information System-Fatigue (PROMIS-Fatigue) Short Form

5.5.2.1. Definition

The PROMIS-Fatigue (Short Form 8a) subscale consists of a static 8-item questionnaire. The PROMIS-Fatigue instrument assesses a range of symptoms from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles. The fatigue short form is universal rather than disease-specific. It assesses fatigue over the past 7 days.

Each question has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For the 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40. Lower scores indicate less fatigue.

"I feel fatigued" ranges from 1=not at all to 5=very much
"I have trouble starting things because I am tired" ranges from 1=not at all to 5=very much
"How run-down did you feel on average?" ranges from 1=not at all to 5=very much
"How fatigued were you on average?" ranges from 1=not at all to 5=very much
"How much were you bothered by your fatigue on average?" ranges from 1=not at all to 5=very much
"To what degree did your fatigue interfere with your physical functioning?" ranges from 1=not at all to 5=very much

"How often did you have to push yourself to get things done because of your fatigue?" ranges from 1=never to 5=always

"How often did you have trouble finishing things because of your fatigue?" ranges from 1=never to 5=always

A score can be approximated if a participant skips a question. Imputation of total score is presented in Section 5.1.3.

The total raw score will be converted into a T-score for each participant based on the table in Appendix 9. The T-score rescales the raw score into a standardized score with a mean of 50 and an SD of 10.

5.5.2.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline timepoint in the double-blind phase will be presented for PROMIS-Fatigue raw score and the T-score by intervention group.

The change from baseline in PROMIS-Fatigue T-score over time will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline PROMIS-Fatigue T-score total score as the covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

5.5.3. Insomnia Severity Index (ISI) Patient Version

5.5.3.1. Definition

The ISI is a 7-item questionnaire assessing the nature, severity, and impact of insomnia. The dimensions evaluated are: sleep onset, sleep maintenance, early morning awakening problems; sleep dissatisfaction; interference of sleep problem with daily functioning; noticeability of sleep problems by others; and distress caused by the sleep difficulties. A 5-point Likert scale (0-4) is used to rate each item. The total score is the sum of the 7 item scores, ranging from 0 to 28. The total score is interpreted as follows: absence of insomnia (0-7); sub-threshold insomnia (8-14); moderate insomnia (15-21); and severe insomnia (22-28).

Negative changes in ISI total score indicate improvement.

5.5.3.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented for patient-reported ISI total score by intervention group.

The change from baseline in ISI total score over time will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline ISI total score as the covariate.

A frequency distribution of the ISI individual item scores over time will be provided by intervention group.

5.5.4. Patient Global Impression of Severity (PGI-S) (Sleep Symptoms)

5.5.4.1. Definition

The PGI-S for sleep symptoms consists of 2 patient-reported items to capture the participant's perceived severity of difficulty falling or staying asleep as well as the problem of not feeling rested the next day. Both items have the recall period of the past 7 days. The 2 dimensions of PGI-S are measured on a 5-point scale: 1=no difficulty of falling asleep or staying asleep (for *severity of difficulty falling and staying asleep*), or I did not have this problem (for *problem of not feeling rested the next day*); 2=mild; 3=moderate; 4=severe; 5=very severe.

Negative changes in PGI-S dimension scores indicate improvement.

5.5.4.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline for each dimension will be presented by intervention group at each postbaseline timepoint in the double-blind phase for observed case data for each dimension.

The difference of change from baseline at each postbaseline timepoint in PGI-S dimension score between seltorexant and quetiapine XR will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline PGI-S dimension score as a covariate.

A frequency distribution of the PGI-S dimension scores over time will be provided by intervention group. The frequencies of PGI-S scores will be plotted over time for the 2 dimensions.

5.5.5. Consensus Sleep Diary

Analysis of Consensus Sleep Diary data will be discussed in an SAP addendum.

5.5.6. Quality of Life in Depression Scale (QLDS)

5.5.6.1. Definition

The Quality of Life in Depression Scale (QLDS) is a disease specific patient-reported outcome (PRO) designed to assess health related quality of life in patients with MDD. (Hunt 1992) The instrument has a recall period of "at the present time", contains 34 items with "true/not true" response options and takes approximately 5 to 10 minutes to complete.

Each statement on the QLDS is given a score of "1" or "0". A score of "1" is indicative of adverse quality of life. The QLDS total score is the sum of the 34 item scores, and ranges from 0 (good quality of life) to 34 (very poor quality of life).

As the measure contains positive as well as negative items, item scores are not always in the same order - that is, not all "true" responses are allocated a score of "1". The scoring of the QLDS items is found in Appendix 10.

Imputation of total score is presented in Section 5.1.3.

5.5.6.1.1. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point will be presented for QLDS total score by intervention group.

The change from baseline in QLDS total score will be analyzed using the same MMRM as described in Section 5.4.2.2.2 for MADRS total score, with the baseline QLDS total score as a covariate.

Least squares mean changes from baseline (+/- SE) will be presented graphically over time.

Frequency distributions of the QLDS individual items will be provided at each assessment time point by intervention group.

5.5.7. European Quality of Life (EuroQol) Group, 5 Dimension, 5-Level questionnaire (EQ-5D-5L)

5.5.7.1. Definition

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome, primarily designed for self-completion by respondents. It is a descriptive system comprised of the following 5 dimensions: Mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each of the 5 dimensions is divided into 5 levels of perceived problems (Level 1 indicating no problem, Level 2 indicating slight problems, Level 3 indicating moderate problems, Level 4 indicating severe problems, and Level 5 indicating extreme problems). (EuroQol Group 2019, EuroQol Group 2023)

Participants select an answer for each of the 5 dimensions considering the response that best matches their health “today”. Individual scores from the 5 dimensions will be used to obtain a weighted health status index as shown below: (i) Scores from each dimension will be combined to obtain a 5L profile score: eg, a score of 1 for each dimension will give a 5L profile score of 11111. Dimension scores will be combined in the following order: Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression. (ii) The value set of the Health Status Index (HSI) for various values of 5L profile scores is published for Canada in the following website: https://www.ncbi.nlm.nih.gov/pubmed/26492214/. (iii) The Canadian value set will be used to get the HSI values for all the countries participating in the study.

A sum score, with a possible range of 0 to 100, is derived as follows: (sum of the scores from the 5 dimensions minus 5) * 5. Negative changes in the sum score indicate improvement.

5.5.7.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point in the double-blind phase will be presented by intervention group for weighted EQ-5D health status index, the EQ-VAS, and the sum score.

Individual dimension responses will also be summarized at each visit with frequency counts and percentage of participants by intervention group.

5.5.8. Clinical Global Impression – Severity (CGI-S) (Depression symptoms)

5.5.8.1. Definition

The CGI-S (Depression) provides an overall clinician-determined summary measure of the severity of the participant's illness that takes into account all available information, including knowledge of the participant's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the participant's ability to function. (Guy 1991) The CGI-S evaluates the severity of psychopathology on a scale of 1 to 7. Considering total clinical experience with the depression population, a participant is assessed on severity of illness at the time of rating according to: 0=not assessed; 1=normal, not at all ill; 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients.

A score of 0 indicates that the participant was not assessed and will be treated as missing. The score will be summarized as recorded.

Negative changes in CGI-S score indicate improvement.

5.5.8.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline will be presented at each post-baseline timepoint in the double-blind phase by intervention group for observed case data.

The change from baseline in CGI-S score at each postbaseline timepoint will be analyzed using the same MMRM model as described for the secondary endpoint MADRS (Section 5.4.2.2.2) with baseline CGI-S score as a covariate.

A frequency distribution of the CGI-S scores over time will be provided by intervention group. The frequencies of CGI-S scores will be plotted over time.

5.5.9. Patient Global Impression of Change (PGI-C) (Depression)

5.5.9.1. Definition

The PGI-C for depression is a single item that will capture the participant's perceptions of improvement or deterioration in depression symptoms compared with when the participant started the study intervention. The change of measured on a 7-point scale: 1=much better; 2=somewhat better; 3=a little better; 4=about the same; 5=a little worse; 6=somewhat worse; 7=much worse.

Higher scores indicate deterioration of symptoms.

5.5.9.2. Analysis Methods

Descriptive statistics of the actual values at each postbaseline time point in the double-blind phase will be presented by intervention group for observed case data.

The difference in PGI-C score at each postbaseline time point in the double-blind phase between seltorexant and quetiapine XR will be analyzed using the same MMRM model as described for the secondary endpoint MADRS without baseline as the covariate.

A frequency distribution of the PGI-C over time will be provided by intervention group.

5.5.10. Sheehan Disability Scale (SDS)

5.5.10.1. Definition

The SDS, a patient-reported outcome measure, is a 5-item questionnaire which has been widely used and accepted for assessment of functional impairment and associated disability. (Leon 1997; Sheehan 1996) The first 3 items assess disruption of (1) work/school work, (2) social life/leisure activities, and (3) family life/home responsibilities using a 0-10 rating scale. The total score is the sum of the first 3 items, ranging from 0 to 30. Higher scores indicate greater impairment. It also has one item on days lost from school or work and one item on days when underproductive. Negative changes in SDS total score indicate improvement.

5.5.10.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline over time will be presented for SDS total score by intervention group.

The change from baseline in SDS total score over time will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline SDS total score as the covariate.

In addition, descriptive statistics of the actual values and the change from baseline, as well as a frequency distribution, will be presented by intervention group for each of the 5 SDS individual item scores at each time point. The percentage of participants who have not worked/studied at all during the past 7 days will be summarized at each time point during the double-blind phase by intervention group.

5.5.11. Ruminative Response Scale (RRS)

5.5.11.1. Definition

The RRS assesses rumination as the process of "compulsively focused attention on the symptoms of one's distress, and on its possible causes and consequences, as opposed to its solutions" as established by Nolen-Hoeksema in 1998. The 22 items of the RRS measure aspects of rumination, brooding and reflective pondering. (Nolen-Hoeksema 2000) A 4-point Likert scale (1=almost never, 2=sometimes, 3=often, 4=almost always) is used to rate each item.

The RRS total score is the sum of the 22 item scores, ranging from 22 to 88. A higher total score indicates higher degree of rumination.

Negative changes in RRS total score indicate improvement.

5.5.11.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline at each postbaseline timepoint will be presented for RRS total score by intervention group.

The change from baseline in RRS total score over time will be analyzed using the same MMRM model as described for the secondary endpoint MADRS with baseline RRS total score as the covariate. Difference of least square means and 2-sided 95% CI will be presented.

5.5.12. Response without Clinically Relevant Weight Gain

The following analysis will be performed on the FAS1CON and FAS2CON analysis sets.

5.5.12.1. Definition

An exploratory endpoint is the proportion of participants who meet the response criteria (≥50% improvement in MADRS total score from baseline) and did not have weight increase of ≥7% at Week 26. Participants who discontinue treatment prior to Week 26 will not be considered meeting the response criteria.

5.5.12.2. Analysis Methods

The number and percentage of participants who achieve response without clinically relevant weight gain will be summarized at Week 26 by intervention group.

5.5.13. Correlation between MADRS total score and PROMIS-SD T-score

Correlation between change in MADRS total score and change in PROMIS-SD T-score at Week 26 will be explored; Pearson and Spearman correlation coefficients will be calculated. Scatter plot of change in scores of the 2 assessment scales at Week 26 will be presented for each intervention group.

5.5.14. Subgroup Analyses

For the subgroups listed in Section 5.7.4, subgroup analyses using CMH will be performed for the proportion of participants who meet the response criteria (≥50% improvement in MADRS total score from baseline) at Week 26. The CMH subgroup analysis will be unstratified. In addition, the CMH test for each subgroup will be stratified by region, age group and baseline rumination level. For the subgroups of country and region, strata will not include region; for age group and age group II, strata will not include age group; for baseline rumination level, strata will not include RRS level.

The analysis results (percent difference and 95% CI based on Wald statistic) of the different subgroups for both unstratified and stratified analysis will be displayed in forest plots.

For the key secondary endpoint of weight change, subgroup analyses using MMRM will be performed for the change in weight at Week 26. The fixed terms in the model will be intervention group, country, age group (adults, elderly), baseline RRS total score levels, time, subgroup, time-by-intervention interaction, intervention-by-subgroup interaction, and time-by-intervention-by-

subgroup interaction, and baseline weight as a covariate. Point estimate of the treatment difference and 2-sided 95% CI will be estimated using appropriate contrasts. Country will not be included in the model where the subgroup of interest is Region, age group will not be included in the model where the subgroup of interest is age group II, and RRS level will not be included in the model where the subgroup of interest is for baseline rumination level.

The analysis results (difference of LS means and 95% CI) of the different subgroups will be displayed in a forest plot.

5.6. Safety Analyses

All safety analyses will be based on the safety analysis set based on planned intervention, unless otherwise specified.

For all continuous safety variables, descriptive statistics by intervention group will include the N, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by intervention group using frequency counts and percentages.

5.6.1. Extent of Exposure

The number and percentage of participants who receive study intervention will be summarized.

Descriptive statistics for duration of study intervention and of the background antidepressant intervention (N, mean, SD, median, and range (minimum, maximum)) will be summarized.

Duration of intervention will be summarized in the following duration categories: [<4 weeks, 4-<6 weeks, 6-<10 weeks, 10-<14 weeks, 14-<18 weeks, 18-<22 weeks, 22-26 weeks, >26 weeks] for each study intervention and presented graphically in a histogram.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1.

Total dose days of intervention is defined as the total number of days that study intervention was administered to the participant (excluding days off study intervention).

The number (%) of participants with a dose adjustment will be summarized by intervention group and visit.

The mode daily dose will be calculated for each subject, based on days on study intervention. The lowest mode dose is selected in case of ties. The frequency of mode daily dose will be presented by study intervention.

The mean daily dose of study intervention is calculated as (sum of total daily dose during the intervention phase)/total dose days of intervention. The final dose is the last non-zero dose received during the double-blind phase. Descriptive statistics (N, mean, SD, median, minimum and maximum) for mean daily dose and final dose will be presented by study intervention.

The analysis will be performed on FAS1, FAS1CON and safety analysis sets.

5.6.1.1. Intervention Compliance

Compliance will be summarized descriptively for each study intervention.

The percent compliance will be categorized and the number and percentage of participants in each category will be summarized by intervention group for study intervention and for the background antidepressant.

Compliance will be calculated for the study intervention as follows:

Compliance (%) = (number of days a participant took expected number of tablets)/(number of days the participant is supposed to be dosed) x 100

Compliance will be calculated for the background antidepressant as follows:

Compliance (%) = (number of days actually dosed)/(days expected to be dosed (includes days where the antidepressant was not taken) x 100

The analysis will be performed on FAS1 and safety analysis sets.

A listing of participants who were impacted by drug supply issues will be provided for the safety analysis set.

5.6.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention through the day of last dose plus 2 days is considered to be treatment emergent. If the event occurs on the day of the initial administration of study intervention, then the event will be assumed to be treatment emergent unless the action taken with study treatment for the event is 'not applicable'. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study intervention based on partial onset date or resolution date. All reported adverse events will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by intervention group.

Summary tables will be provided for treatment-emergent adverse events by phase:

- AEs (all AEs, and AEs with incidence of at least 5% in any intervention group)
- Serious AEs (SAEs)
- AEs leading to discontinuation of study intervention
- AEs by severity
- AEs by relationship to study intervention (DB phase only)

- AEs of special interest

In addition to the summary tables, listings will be provided for participants who had:

- SAEs
- AEs leading to discontinuation of study intervention
- AEs of special interest

A listing of participants who died will be provided.

5.6.3. Additional Safety Assessments

5.6.3.1. Clinical Laboratory Tests

Descriptive statistics will be presented for all chemistry, hematology, and urinalysis laboratory tests at scheduled time points in the double-blind phase.

Change from baseline to all postbaseline visits will be summarized for chemistry, hematology, and urinalysis tests and displayed by study intervention group.

Clinical laboratory test values will be considered "treatment-emergent markedly abnormal" (TEMA) using the criteria defined by the sponsor listed in Appendix 11.

- If the postbaseline value is above the upper limit of the markedly abnormal criteria and the baseline value is equal to or below the upper limit, then the postbaseline marked abnormality will be considered TE. The same applies to the postbaseline value being below the lower limit of the markedly abnormal criteria with the baseline value being equal to or above the lower limit of the markedly abnormal criteria.
- If the baseline value is missing, a postbaseline marked abnormality will always be considered as TE.

The number and percentage of participants with treatment-emergent markedly abnormal values will be presented by study intervention group.

The incidence of participants with treatment-emergent ALT values >3*upper normal limit (ULN) or AST value > 3*ULN will be presented for the double-blind phase. Additionally, incidence of treatment-emergent hepatic toxicity (suspected Hy's Law (US Dept Health 2009) cases) defined as (ALT values >3*ULN or AST values > 3*ULN) AND total bilirubin values >2*ULN will be presented for the double-blind phase. Similar to the markedly abnormal analysis, only participants with baseline values that do not meet criteria or if baseline value is missing will be eligible for these analyses.

A listing of participants with treatment-emergent markedly abnormal laboratory values will be provided. A listing of participants with ALT > 3* ULN or AST values > 3*ULN) and participants with hepatic toxicity (suspected Hy's Law cases) will be provided.

In addition, the number of participants with the following shifts in chemistry laboratory values from baseline to the maximum postbaseline time point will be presented for double-blind phase:

Glucose:

- from <100 mg/dL to [≥100 mg/dL - <126 mg/dL] (normal to borderline)
- from [≥100 mg/dL - <126 mg/dL] to ≥126 mg/dL (borderline to high)
- from <100 mg/dL to ≥126 mg/dL (normal to high)

Triglycerides:

- from <150 mg/dL to ≥200 mg/dL (normal to high/very high)
- from <150 mg/dL to ≥500 mg/dL (normal to very high)
- from [≥150 mg/dL - <200 mg/dL] to ≥200 mg/dL (borderline to high/very high)
- from [≥150 mg/dL - <200 mg/dL] to ≥500 mg/dL (borderline to very high)
- from [≥200 mg/dL - <500 mg/dL] to ≥500 mg/dL (high to very high)

Total Cholesterol

- from <200 mg/dL to ≥200 mg/dL (normal to borderline/high)
- from <200 mg/dL to ≥240 mg/dL (normal to high)
- from <200 mg/dL to [≥200 mg/dL - <240 mg/dL] (normal to borderline)
- from [≥200 mg/dL - <240 mg/dL] to ≥240 mg/dL (borderline to high)

HDL Cholesterol

- from ≥40 mg/dL to <40 mg/dL (normal to low)

LDL Cholesterol

- from <100 mg/dL to ≥160 mg/dL (normal to high).

Descriptive statistics for metabolic and lipid profile laboratory parameters (Glucose, HbA1c, Triglycerides, Total Cholesterol, HDL and LDL Cholesterol) along with 95% confidence intervals for change from baseline values will be summarized at scheduled time points in the double-blind phase.

5.6.3.1.1. Homeostatic Assessment (HOMA) Modeling

Insulin resistance and beta-cell function based upon fasting glucose and insulin using the homeostatic assessment (HOMA) (Matthews, 1985) model will be assessed. Two variables, HOMA IR (insulin resistance) and HOMA-%B (beta-cell function) will be derived. The relationship between glucose and insulin secretion, mathematically approximated using a simple nonlinear solution, is given below:

$$HOMA\ Insulin\ Resistance\ (IR) = \frac{FI}{22.5e^{-lnFG}}$$

$$HOMA\ Beta\ Function\ (B) = \frac{20\ x\ FI}{[FG - 3.5]}$$

where FG = fasting glucose (mmol/L); FI = fasting insulin (mU/L); Insulin:
1 µIU/mL = 6.945 pmol/L. HOMA IR and HOMA-%B will not be derived if FG is <=3.5.

The descriptive statistics for HOMA-IR and HOMA-%B at baseline (DB) and end point (DB) for the safety analysis set will include the following:

Geometric mean (GM) = exp(mean(logs));

GM mean ± 1 SD = (exp(mean(logs) - 1 SD(logs)), exp(mean(logs) + 1 SD(logs)));

where logs indicates the natural logarithm of the HOMA values.

In addition, descriptive statistics for HOMA-IR and HOMA-%B along with 95% confidence intervals for change from baseline values will be summarized at scheduled time points in the double-blind phase.

5.6.3.2. Vital Signs and Physical Examination Findings

Continuous vital sign parameters including temperature, weight, waist circumference, pulse, blood pressure (systolic and diastolic), and Body Mass Index (BMI) will be summarized at each assessment time point. Body Mass Index will be calculated as weight (kg)/(height (m))^2, at each time point that body weight is measured. The height measurement collected at screening will be used in the calculation. Change from baseline will be summarized for the double-blind phase. Descriptive statistics (mean, standard deviation, median, minimum and maximum) will be presented. For weight, percent changes from baseline will be presented.

Abnormality criteria (based on criteria defined below) will be applied to postbaseline values in the double-blind phase.

Postbaseline values will be considered treatment-emergent if they meet both value and change criteria in the table below.

For criteria that do not include an increase or decrease from baseline:

- Treatment-emergent will be concluded if the postbaseline value is above the upper limit and the baseline value is below the upper limit (eg, Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (eg, Normal or High).
- If the baseline value is missing, a postbaseline abnormality will always be considered as TE.

Incidence of treatment-emergent clinically important vital signs during intervention, as defined in Table 5, will be summarized for participants who had a baseline assessment and at least 1 postbaseline assessment for that vital sign. A listing of participants with treatment-emergent clinically important abnormalities in vital signs will be presented, along with a listing of all vital sign measurements.

Table 5: Clinically Important Abnormalities in Vital Signs

Vital Sign	Abnormal Category	Criteria
Pulse	Abnormally high	≥100 bpm and with ≥15 bpm increase from baseline
	Abnormally low	≤50 bpm and with ≥15 bpm decrease from baseline
Systolic blood pressure	Abnormally high	≥180 mm Hg and with ≥20 mm Hg increase from baseline
	Abnormally low	≤90 mm Hg and with ≥20 mm Hg decrease from baseline
Diastolic blood pressure	Abnormally high	≥105 mm Hg and with ≥15 mm Hg increase from baseline
	Abnormally low	≤50 mm Hg and with ≥15 mm Hg decrease from baseline
Temperature	Abnormally high	>37.5ºC
	Abnormally low	<35.5ºC
Weight	Abnormally high	increase ≥7% from baseline
	Abnormally low	decrease ≥7% from baseline

5.6.3.3. Electrocardiogram

The ECG parameters that will be analyzed are heart rate, PR interval, QRS interval, QT interval, and corrected QT (QTc) interval using the following correction methods: Bazett's formula (QTcB), Fridericia's formula (QTcF).

Bazett's formula: QTcB (msec) = QT (msec) / (RR (msec)/1000)$^{1/2}$; if RR is missing, use QT (msec) * (HR(bpm)/60)$^{1/2}$;

Fridericia's formula: QTcF (msec) = QT (msec) / (RR (msec)/1000)$^{1/3}$; if RR is missing, use QT (msec) * (HR(bpm)/60)$^{1/3}$;

Baseline ECG is defined as the average of all ECG results collected up to and including the day of the first dose of study intervention.

The number and percentage of participants with QTc interval increases from average baseline to the maximum postbaseline value will be summarized at each scheduled time point. Refer to the following Table 6 for summary categories.

Table 6: Criteria for Abnormal QTc Values and Changes From Baseline

QTc value (msec)	Normal QTc	≤450 for male, ≤470 for female
		>450 to ≤480 for male, >470 to ≤480 for female
		>480 to ≤500
		>500
Clinically significant QTc (msec)	No	≤500
	Yes	>500
QTc change from baseline (msec)	No concern	≤30
	Concern	>30-60
	Clear concern	>60

Descriptive statistics of ECG parameters and change from average baseline will be summarized at each scheduled time point.

Abnormality criteria (based on criteria defined below) will be applied to average baseline and postbaseline values.

Postbaseline abnormalities will be compared with their corresponding average baseline result:

- Treatment-emergent will be concluded if the postbaseline value is above the upper limit and the average baseline value is below the upper limit (eg, Normal or Low). The same applies to the postbaseline value being below the lower limit with the average baseline value being above the lower limit (eg, Normal or High).
- If the average baseline value is missing, a postbaseline abnormality will always be considered as treatment-emergent.

The number and percentage of participants with treatment-emergent ECG values outside predefined limits will be presented by study intervention over time for the double-blind phase of the study:

Table 7: Abnormal Limits for ECG Parameters

ECG Parameter	Abnormally low	Abnormally high
Heart Rate (bpm)	≤ 50 bpm	≥100 bpm
PR interval (msec)	≤ 120 msec	≥ 200 msec
QRS interval (msec)	≤ 60 msec	≥120 msec
QT interval (msec)	≤ 200 msec	≥500 msec
RR interval (msec)	< 600 msec	> 1200 msec

The interpretation of the ECGs as determined by a central reader will be displayed by the number and percentage of participants meeting the normality criteria. The interpretation will be summarized over time.

A listing of clinically relevant ECG abnormalities will also be provided.

5.6.3.4. Other Safety Parameters

5.6.3.4.1. Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a low-burden measure of the spectrum of suicidal ideation and behavior that was developed to assess severity and track suicidal events through any intervention (Posner 2007). It is a clinical interview providing a summary of both suicidal ideation and behavior that can be administered during any evaluation or risk assessment to identify the level and type of suicidality present. The C-SSRS has been used frequently in clinical studies, and is a validated, standard measure for suicidal ideation assessment. Using the C-SSRS, the outcomes will be categorized using the scoring for the 11 categories:

Suicidal Ideation (1-5)	
1	Wish to be dead
2	Non-specific active suicidal thoughts
3	Active suicidal ideation with any methods (not plan) without intent to act
4	Active suicidal ideation with some intent to act, without specific plan
5	Active suicidal ideation with specific plan and intent
Suicidal Behavior (6-10)	
6	Preparatory acts or behavior
7	Aborted attempt
8	Interrupted attempt
9	Actual attempt
10	Suicide
Non-suicidal self-injurious behavior (11)	
11	Non-suicidal self-injurious behavior

At each time point, an event of suicidal ideation or behavior will be assigned a score of 1 to 10 based on the maximum response for the C-SSRS at that visit. If no events qualify for a score of 1 to 10, a score of 0 will be assigned (0="no suicidal ideation or behavior that can be assessed on the basis of C-SSRS"). A participant with an event of non-suicidal self-injurious behavior only will not be considered as having suicidal ideation or behavior, therefore, a score of 0 will be assigned. However, an additional score of 11 will be assigned to summarize any participants with an event of non-suicidal self-injurious behavior.

Shifts from the baseline visit to the maximum postbaseline score pertaining to suicidal ideation or suicidal behavior (i.e., scores 1 to 10) will be summarized by intervention for each phase.

The maximum score (of scores 1 to 10) assigned to each participant will be grouped into one of three broad categories: No suicidal ideation or behavior (0), Suicidal ideation (1-5), Suicidal behavior (6-10). Shifts from the baseline visit to the maximum postbaseline category will be summarized by intervention for each phase.

A frequency distribution of the scores for the 11 categories will be provided at each time point for each phase.

5.6.3.4.2. Arizona Sexual Experiences Scale (ASEX)

The ASEX is a patient-reported 5-item rating scale that quantifies sex drive, arousal, vaginal lubrication/penile erection, ability to reach orgasm, and satisfaction from orgasm. The scale has shown satisfactory reliability and validity. (McGahuey 2000)

Each of the 5 items is rated on a 6-point Likert scale, ranging from 1 to 6. The total score is the sum of the 5 items. Higher scores indicate more sexual dysfunction. If any item of the scale is missing at a visit, the total score for that scale at that visit will be left blank.

For each of the 5 items, a frequency distribution will be provided by intervention group and gender at each time point. In addition, for each of the 5 items, a frequency distribution will be provided

by intervention group, combining the responses for the genders. For this analysis, "vaginal lubrication/penile erection" (question 3) will be summarized as one question. The ASEX total score at each time point and the change from baseline will be summarized with descriptive statistics by intervention group.

The number and percentage of participants who have ASEX total score 19 or greater, or a score of 5 or greater on any item, or a score of 4 or greater on any 3 items, reflecting sexual dysfunction, will be summarized at each time point by intervention group.

5.6.3.4.3. Cognitive Assessment (Revere.D)

The Revere.D computerized cognitive test battery is self-administered by the participant on an iPad and includes 8 tests covering a broad range of cognitive functions. The software provides immediate scoring of responses. The completion of these cognitive tests requires approximately 25 minutes.

Test	Cognitive Domain
Symbol Sorting Test (SST)	Executive function
Digit Span Forward Test (DSFT)	Attention, working memory
Digit Span Backward Test (DSBT)	Attention, working memory
Word List Recall Test (WLRT)	Verbal episodic memory
Visuospatial Block Recall Test (VBRT)	Visuospatial memory
Symbol Digit Matching Test (SDMaT)	Divided attention, processing speed
Trail Making Test - Part B (TMT-B)	Processing speed, sequencing / executive function
Block Maze Test (BMT)	Executive function, visuospatial learning / working memory, error processing

- Symbol Sorting Test (SST): The participant is asked to sort symbols, based on symbol features including color, shape, and number. The task requires the participant to decipher sorting rules based on feedback whether each sort was correct or incorrect. Rules change unannounced at various points during the task. Responses are scored for Categories Achieved and Percent Perseverative Errors.
- Digit Span Forward Test (DSFT) and Digit Span Backward Test (DSBT): The participant is asked to repeat a sequence of digits in the order presented (DSFT) or in reverse order (DSBT). Scores are determined based on longest span of correct responses.
- Word List Recall Test (WLRT): The participant is asked to verbally recall a list of 15 words presented aloud. After 5 trials with the same 15-word list, a second list is presented for recall, after which the participant is asked to recall words from the original word list. After 20 minutes, during which the participant completes other tests in the Revere.D battery, the participant is asked to again recall the original word list. Scores include the number of words recalled across the 5 trials of the original word list (total learning score) and on the delayed recall.
- Visuospatial Block Recall Test (VBRT): The participant is asked to recall the location of colored tiles within an onscreen grid; during each trial the colored tiles are briefly

illuminated on-screen. The test progresses through more complex tile patterns. The score is the number of correctly recalled patterns.

- Symbol Digit Matching Test (SDMaT): The participant is asked to pair 9 abstract symbols with specific numbers (1 to 9) using a predefined coding key as fast as possible using a keypad. The score is the number of correct matches within 90 seconds.
- Trail Making Test - Part B (TMT-B): The participant is asked to connect a set of consecutively numbered and lettered circles, alternating between numbers and letters (ie, 1 A 2 B). The score is the time to completion.
- Block Maze Test (BMT): The participant is asked over 3 trials to learn a path across a tile grid from a starting position to a target position. After each "step" made in the maze grid, feedback is provided indicating whether the step is correct or incorrect. The score is the total number of errors made in the 3 trials.

Descriptive statistics of the actual values and the change from baseline over time will be presented for each cognitive assessment domain by intervention group. The cognitive assessment scores will be plotted over time.

In addition, Reliable Change Index (RCI) values will be derived. Refer to Appendix 13 for details.

Percentage of RCI values < -1.96 and < -1.65 will be presented over time for each cognitive assessment domain by intervention group.

5.6.3.4.4. Physician Withdrawal Checklist - 20 items (PWC-20)

The PWC-20 is a reliable and sensitive instrument for the assessment of discontinuation symptoms. (Rickels 2008) The PWC-20 is a simple and accurate method used to assess potential withdrawal symptoms following cessation of study intervention.

The proportion of participants with withdrawal symptoms at end of the double-blind phase and follow-up visits will be presented by intervention group.

In addition, symptoms at follow-up will be compared to the end of therapy visit and will be summarized using the following categories: new or worsened symptoms, symptoms present and unchanged, no symptoms, and improved.

Additionally, PWC-20 total score will be computed, and summary statistics of the total score at each visit will be provided.

Bar chart of the distribution of PWC scores over time will be presented for each item.

5.6.3.4.5. Menstrual Cycle Tracking

Menstrual cycle in premenopausal women will be tracked during the double-blind and follow-up phases. Participants who are experiencing menses during the screening evaluation belonging to the safety analysis set will be included in the analysis.

Descriptive summary statistics (n, mean, SD, median, minimum, and maximum) of the duration of menses and the length of menstrual cycle will be provided for the duration of the double-blind phase. This analysis will be repeated for the following 2 sub-groups: women who use contraception, and those who don't use contraception. Duration of menses in the follow-up phase will be summarized for the follow-up analysis set.

5.7. Other Analyses

5.7.1. Pharmacokinetics

Blood samples were collected for participants who discontinued study drug for an AE, had an AESI or had an SAE if the sample could be obtained within 15 hours of the last study drug administration. Because a limited number of samples were collected, analysis will not be performed and no listing will be provided.

5.7.2. Biomarkers

Analysis of biomarker data will be discussed in a separate SAP.

5.7.3. Health Economics

5.7.3.1. Healthcare Resource Use Questionnaire (HRUQ)

Medical resource utilization data, associated with medical encounters, will be collected using the HRUQ during the double-blind phase of the study.

The number and percentage of participants who had at least one outpatient visit to a healthcare professional, related to their depression, in the period since the most recent assessment preceding each time point, will be summarized for each type of healthcare professional and for any healthcare professional by time point.

The number and percentage of participants who had at least one emergency room visit without hospitalization, related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The number and percentage of participants who had at least one day clinic or day treatment program (no overnight stay), related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The number and percentage of participants who had at least one hospitalization (minimum one overnight stay), related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The total number of outpatient visits to each type of healthcare professional, related to depression, and the total number of outpatient visits to any healthcare professional, related to depression, in the period since the most recent assessment preceding each time point, will be summarized with descriptive statistics at each time point.

Likewise, the total number of emergency room visits without hospitalization, related to depression, the total number of day clinic or day treatment programs (no overnight stay), related to depression, and the total number of hospitalizations (minimum one overnight stay), related to depression, will be summarized with descriptive statistics at each time point.

All summaries will be presented by intervention group.

5.7.4. Definition of Subgroups

Subgroup analyses of the primary (FAS1CON and FAS2CON analysis sets) and key secondary (FAS1 and FAS2 analysis sets) endpoints will be performed for the following subgroups:

Subgroup	**Definition**
Sex	Male Female Other
Race	Asian Black White Other
Ethnicity	Hispanic or Latino Not Hispanic or Latino
Country	Argentina, Belgium, Bulgaria, Canada, Czech Republic, Estonia, Latvia, Lithuania, Malaysia, Poland, Russia, Serbia, Slovakia, UK, Ukraine and USA
Region	USA European Union+UK Rest of World
Age Group	Adult (<65 years) Elderly (≥ 65 years)
Age Group II	18-34 years 35-54 years 55-64 years ≥ 65 years
Baseline BMI	underweight <18.5 kg/m^2 normal 18.5-<25 kg/m^2 overweight 25-<30 kg/m^2 obese >=30 kg/m^2
Number of major depressive episodes	< 3 ≥ 3
Number of failed prior antidepressants	1 2 3
Current antidepressant type	SSRI SNRI
Prior use of benzodiazepines	Yes No
Baseline RRS Score based on IWRS	<54 ≥ 54

5.8. Interim Analyses

One interim analysis of the primary endpoint was planned for this study if one of two short-term studies were stopped for futility. The aim of the IA was to stop the study for futility or initiate a new short-term study. Since Protocol Amendment 5, a decision was made to not perform this IA.

5.8.1. Independent Data Monitoring Committee (IDMC)

An independent data monitoring committee will monitor the SAEs on a bi-monthly basis. Additionally, the IDMC will meet periodically to review safety data and meet once to evaluate the unblinded efficacy data at the IA and make recommendation on whether to declare futility. Details about the membership of the IDMC and its roles and responsibilities are discussed in a separate IDMC charter.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AE	adverse event
AESI	adverse event(s) of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ASEX	Arizona Sexual Experiences Scale
AST	aspartate aminotransferase
ATC	anatomic and therapeutic class
BMI	body mass index
C-SSRS	Columbia Suicide Severity Rating Scale
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
CRF	case report form
CSD	Consensus Sleep Diary
CV	coefficient of variation
DB	double-blind
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ECG	electrocardiogram
eDC	electronic data capture
EQ-5D-5L	European Quality of Life (EuroQol) Group, 5 Dimension, 5-Level
EQ-VAS	European Quality of Life-Visual analog Scale
EU	European Union
FAS1	full analysis set 1
FAS2	full analysis set 2
FDA	Food and Drug Administration
FWER	familywise error rate
HOMA	homeostatic assessment
HRUQ	Healthcare Resource Use Questionnaire
HSI	health status index
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
IQ	interquartile
ISI	Insomnia Severity Index
IWRS	interactive web response system
LOCF	last observation carried forward
MADRS	Montgomery-Asberg Depression Rating Scale
MADRS-WOSI	Montgomery-Asberg Depression Rating Scale WOSI (Without Sleep Item)
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MDD	major depressive disorder
MDDIS	major depressive disorder with insomnia symptoms
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	Mixed-Effect Model for Repeated Measures
MPV	Major protocol violations
NIH	National Institutes of Health
PD	pharmacodynamic(s)
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression – Severity
PHQ-9	Patient Health Questionnaire - 9 Item
PK	pharmacokinetic(s)
PROMIS-SD	Patient Reported Outcome Measurement Information System-Sleep Disturbance
PROMIS-Fatigue	Patient Reported Outcome Measurement Information System-Fatiguq
PWC-20	Physician Withdrawal Checklist – 20 items

QLDS	Quality of Life in Depression Scale
RCI	Reliable Change Index
RRS	Ruminative Response Scale
s-nNAW	subjective number of nighttime awakenings
SAE	serious adverse event
SAP	Statistical Analysis Plan
SCID-CT	Structured Clinical Interview for DSM-5-Clinical Trial Version
SD	standard deviation
SDS	Sheehan Disability Scale
SNRI	serotonin and norepinephrine reuptake inhibitor
sQUAL	subjective quality of sleep
SSRI	selective serotonin reuptake inhibitor
sSOL	subjective sleep onset latency
sTST	subjective total sleep time
sWASO	subjective wake after sleep onset
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent markedly abnormal
ULN	upper limit of normal
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

6.2. Appendix 2 Changes to Protocol-Planned Analyses

The primary endpoint of Response Rate at Week 26 under estimand 1 was planned as a stratified Cochran-Mantel-Haenzsel (CMH) test adjusted for country, age group and baseline rumination level to test the proportion difference between treatment groups. Mainly due to sparse enrollment in some countries, there are many stratums which will potentially not have representation from both treatment groups and hence will have no contribution to the stratified CMH test statistic. Therefore, the analysis is changed to replace country with geographic region (USA, European Union+UK, Rest of world). A sensitivity analysis is added as described in the protocol using the stratification level of country instead of region.

The primary endpoint under estimand 2 was planned as a stratified CMH test adjusted for country, age group, baseline rumination level and MADRS level. Because a very low percentage of participants are enrolled in the low MADRS level, the analysis is changed to match the analysis under estimand 1. A sensitivity analysis is added as described in the protocol using the stratification levels of country and baseline MADRS level.

6.3. Appendix 3 Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized and listed by intervention group and overall. In addition, the distribution of participants by region, country, and site ID will be presented unless otherwise noted.

Table 8 presents a list of the demographic variables that will be summarized by intervention group and overall for the FAS1, FAS1CON, FAS2 and FAS2CON analysis sets.

Table 8: Demographic Variables

Continuous Variables	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Waist circumference (cm)	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Age Group (Adults, Elderly)	
Age Group II (18-34 years, 35-54 years, 55-64 years, and ≥65 years)	
Sex (male, female, unknown, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino, Not Reported, Unknown)	
BMI (underweight <18.5 kg/m^2, normal 18.5-<25 kg/m^2, overweight 25-<30 kg/m^2, obese >=30 kg/m^2)	
Country	
Region	

[a]If multiple race categories are indicated, the Race is recorded as 'Multiple'

Table 9 presents a list of the baseline disease characteristics variables that will be summarized by intervention group and overall for the FAS1, FAS1CON, FAS2 and FAS2CON analysis sets.

Table 9: Baseline Disease Characteristics

Continuous Variables	**Summary Type**
Age (years) when diagnosed with MDD	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Duration (weeks) of current depressive episode	
Baseline MADRS total score	
Baseline CGI-S score	
Baseline patient-rated ISI total score per ePRO/eDC	
Baseline clinician-rated ISI total score per eDC	
Baseline RRS total score per eDC	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Baseline CGI-S (Depression) score (Normal (not at all ill), Borderline mentally ill, Mildly ill, Moderately ill, Markedly ill, Severely ill, Among the most extremely ill patients)	
Baseline RRS total score per IWRS (<54, ≥54)	
Baseline RRS total score per eDC ((<54, ≥54)	
Current antidepressant type (SSRI, SNRI)	
Number of failed prior antidepressants (1, 2, 3 or more)	
SCID-CT DSM-5 specifiers for MDD	
Prior medication use of benzodiazepines	
Number of major depressive episodes to date, including current episode (<3, ≥ 3)	
Family history of alcohol abuse (Yes, No)	

Family history of anxiety disorder (Yes, No)	
Family history of bipolar disorder (Yes, No)	
Family history of depression (Yes, No)	
Family history of schizophrenia (Yes, No)	
Family history of substance abuse (Yes, No)	

6.4. Appendix 4 Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations will be summarized by category for the safety analysis set.

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Other

Number of participants not meeting inclusion criteria, or meeting exclusion criteria will be summarized by study intervention group for the safety analysis set.

Additional analyses of protocol deviation data for assessing and mitigating the impact of COVID-19 on study outcome are presented in Appendix 12.

6.5. Appendix 5 Prior and Concomitant Medications

Prior medications and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of first dose (partial or complete) of study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continue on after the first dose of study intervention.

Summaries of concomitant medications will be presented by ATC level 2 term and base preferred term, intervention group, and study phase: double-blind phase/follow-up phase. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication. The following data will be summarized separately: concomitant medications other than antidepressants taken during each phase, and antidepressants taken during each phase. In addition, concomitant medications of special interest will be summarized. See Appendix 7 for list of categories of medications of special interest.

Prior antidepressant medications, prior medications other than antidepressants, and prior medications of special interest will be summarized by intervention group and ATC level 2 term, ATC level 4 term and base preferred term.

6.6. Appendix 6 Adverse Events of Special Interest

Adverse events of special interest are defined as follows:

AE Special Interest Category	Event Type
Cataplexy	Cataplexy
Sleep paralysis	Sleep paralysis
Complex, sleep-related behaviors/parasomnias	Confusional arousal
	Somnambulism
	Sleep terrors
	Bruxism
	Sleep sex
	Sleep-related eating disorder
	Sleep behavior disorder
	Catathrenia
Fall	Fall
Motor vehicle accident	Motor vehicle accident

6.7. Appendix 7 Medications of Special Interest

Categories of medications of special interest are defined as follows:

Medications of Special Interest Category
Antidepressants
Benzodiazepines
Hypnotics/sedatives including z-drugs
Antipsychotics

6.8. Appendix 8 Conversion of Raw Score to T-Score for PROMIS-SD

Sleep Disturbance 8a *Short Form Conversion Table*		
Raw Score	**T-Score**	**SE***
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.8	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE= Standard Error on T-score metric

Adult version

6.9. Appendix 9 Conversion of Raw Score to T-Score for PROMIS-Fatigue

Fatigue 8a *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	33.1	4.8
9	38.5	2.7
10	41.0	2.2
11	42.8	2.0
12	44.3	1.9
13	45.6	1.8
14	46.9	1.8
15	48.1	1.8
16	49.2	1.8
17	50.4	1.8
18	51.5	1.7
19	52.5	1.7
20	53.6	1.7
21	54.6	1.7
22	55.6	1.7
23	56.6	1.7
24	57.5	1.7
25	58.5	1.7
26	59.4	1.7
27	60.4	1.7
28	61.3	1.7
29	62.3	1.7
30	63.3	1.7
31	64.3	1.7
32	65.3	1.7
33	66.4	1.7
34	67.5	1.7
35	68.6	1.7
36	69.8	1.8
37	71.0	1.8
38	72.4	2.0
39	74.2	2.4
40	77.8	3.7

*SE = Standard Error

6.10. Appendix 10 Scoring of QLDS Items

I just want time to pass	True	1
	Not true	0
I feel hopeful about the future	True	0
	Not true	1
I find it hard to hold a conversation	True	1
	Not true	0
I like to know what is going on in the world	True	0
	Not true	1
I feel as if my life is wasting away	True	1
	Not true	0
I feel as if I am not in control of my life	True	1
	Not true	0
I am reluctant to leave the house	True	1
	Not true	0
On the whole, I enjoy the things I do	True	0
	Not true	1
I have lost all pleasure in life	True	1
	Not true	0
I feel as if I have nothing to offer anyone	True	1
	Not true	0
I turn away from people I care about	True	1
	Not true	0
I take good care of myself	True	0
	Not true	1
I am able to think about the future	True	0
	Not true	1
I just want to hide away	True	1
	Not true	0
I look forward to things	True	0
	Not true	1
I've forgotten what it's like to enjoy myself	True	1
	Not true	0
I can't be bothered with my friends	True	1
	Not true	0
I can cope easily with everyday tasks	True	0
	Not true	1

I cut myself off from other people	True	1
	Not true	0
It is difficult for me to make even simple decisions	True	1
	Not true	0
I feel as if I am a burden to people	True	1
	Not true	0
Most of the time I just sit and stare into space	True	1
	Not true	0
I can't face anyone	True	1
	Not true	0
I shut everything out	True	1
	Not true	0
I'm neglecting my appearance	True	1
	Not true	0
I can see the funny side of things	True	0
	Not true	1
I don't take in what people say to me	True	1
	Not true	0
I feel as if I'm letting everyone down	True	1
	Not true	0
I dread each coming day	True	1
	Not true	0
I enjoy my food	True	0
	Not true	1
I avoid people if I can	True	1
	Not true	0
I am reluctant to answer the door or the telephone	True	1
	Not true	0
My life has no meaning	True	1
	Not true	0
I am able to cope with everyday problems	True	0
	Not true	1

6.11. Appendix 11 Criteria for Treatment-emergent Markedly Abnormal Laboratory Values

Laboratory Parameter	Unit	Low	High
Clinical Chemistry			
Albumin	g/dL	2.4	6.0
Albumin	g/L	24	60
Alkaline phosphatase	U/L	N/A	250
Alanine transaminase (SGPT)	U/L	N/A	200
Aspartate transaminase (SGOT)	U/L	N/A	250
Bicarbonate	mEq/L	15.1	34.9
Bicarbonate	mmol/L	15.1	34.9
Bilirubin (direct)	mg/dL	N/A	3.0
Bilirubin (direct)	µmol/L	N/A	51.3
Bilirubin (total)	mg/dL	N/A	3.0
Bilirubin (total)	µmol/L	N/A	51.3
Blood urea nitrogen	mg/dL	N/A	50
Blood urea nitrogen	mmol/L	N/A	17.9
Calcium	mg/dL	6	12
Calcium	mmol/L	1.497	2.994
Chloride	mEq/L or mmol/L	94	112
Cholesterol	mg/dL	N/A	300
Cholesterol	mmol/L	N/A	7.758
Creatine kinase	U/L	N/A	990
Creatinine	mg/dL	N/A	3
Creatinine	µmol/L	N/A	265.2
Gamma glutamyl transferase	U/L	N/A	300
Glucose Plasma	mg/dL	40	300
Glucose Plasma	mmol/L	2.204	16.653
Hemoglobin A1c	fraction of 1	0.04	0.08
High-density lipoprotein cholesterol (HDL)	mg/dL	35	N/A
High-density lipoprotein cholesterol (HDL)	mmol/L	0.905	N/A
Low-density lipoprotein cholesterol (LDL)	mg/dL	89	160
Low-density lipoprotein cholesterol (LDL)	mmol/L	2.3015	4.1376
Phosphate	mg/dL	2.2	8.1
Phosphate	mmol/L	0.71038	2.61549
Potassium	mmol/L	3.0	5.8
Sodium	mEq/L	125	155
Sodium	mmol/L	125	155
Total protein	g/L	50	N/A
Triglycerides	mg/dL	N/A	500
Triglycerides	mmol/L	N/A	5.645
Uric acid	mg/dL	1.5	10
Uric acid	µmol/L	89.22	594.8

Laboratory Parameter	Unit	Low	High
Hematology			
Hematocrit - female	%	0.28	0.50
- male	%	0.24	0.55
Hemoglobin	g/dL	8	19
Hemoglobin	g/L	80	190
Neutrophils	%	30	90
Monocytes	%	N/A	20
Eosinophils	%	N/A	10
Basophils	%	N/A	6
Lymphocytes	%	10	60
Reticulocytes	%	0.5	1.5
Platelet count	10^9/L; giga/L	100	600
Red blood cell (RBC) count - female	10^{12}/L; tera/L	3.0	5.5
- male	10^{12}/L; tera/L	3.0	6.4
White blood cell (WBC) count	10^9/L; giga/L	2.5	15.0
Urinalysis			
Urine pH		N/A	6.5
Urine specific gravity		< 1.001	> 1.035

Note: Values should be flagged as markedly abnormally low if the value is less than the value indicated in the "Low" column. Likewise, values should be flagged as markedly abnormally high if the value is greater than the value indicated in the "High" column.
Note: The same limits apply to both males and females unless gender is indicated.
N/A = Not applicable.

6.12. Appendix 12 Analyses for Assessing and Mitigating the Impact of COVID-19 on Study Outcome

The following measures will be taken to handle the impact of COVID-19 on study outcome:

1. Listing and summary of treatment discontinuation and study discontinuation including reasons due to COVID-19 will be presented
2. Protocol deviation related to COVID-19 including missing visits and remote visits due to COVID will be summarized; corresponding listing will be provided
3. Additional analyses to assess the potential impact of COVID-19 in primary endpoint (MADRS response rate at Week 26, with early treatment discontinuation considered as non-response).
    a. Provide proportion of MADRS response rate at Week 26 by types of collection
    b. Sensitivity analyses will be provided by including type of collection in the primary model for MADRS response at Week 26. If the number of non-site visits is small, then the sensitivity analysis may not be conducted as the results would not be expected to differ from the main analyses.
4. Additional analyses to assess the potential impact of COVID-19 in MADRS.
    a. Mean summaries over time by types of collection (remote, home, site, etc).
    b. Sensitivity analyses will be provided by including types of collection in the primary model for MADRS.

6.13. Appendix 13 Analyses Methods for Revere.D Data

For each test in the Revere,D computerized cognitive test battery, Reliable Change Index (RCI) values will be derived using population norms (see Table 10). The equation for calculating RCI is below:

$$RCI = \frac{(x_2 - x_1)}{\sigma_1 * \sqrt{2(1-r)}}$$

Where x_1 = the baseline score, x_2 = the score at a post-baseline visit, σ_1 = the standard deviation of the normative sample, and r = the normative test-retest reliability (see Table 11).

For tests for which a lower score indicates better performance (Trail Making Test, Block Maze Test), the direction of the RCI values are reversed so that positive values indicate improvement for all tests.

Table 10: Population norms for Revere.D Cognitive Tests

	SST NCC	SST % PE	DSFT Raw Score	DSBT Raw Score	WLRT-Rater LS	WLRT-Rater DRS	VBRT Num DCR	SDMaT TS	TMT-B TTC	BMT Num Errors
Mean	4.9	15.9	8.7	6.9	47.6	10.0	8.3	37.6	78.0	29.2
SD	1.60	7.63	2.42	2.95	10.15	3.35	2.29	13.71	59.39	7.05

Abbreviation: BMT = Block Maze Test; CI = Confidence Interval; DCR = Designs Correctly Recalled; DRS = Delayed Recall Score; DSBT = Digit Span Backward Test; DSFT = Digit Span Forward Test; ICC = Intraclass Correlation Coefficient; LS = Learning Score; NCC = Number of Categories Completed; PE = Perseverative Errors; R=Revere.D; SD = standard deviation; SDMaT = Symbol Digit Matching Test; SST = Symbol Sorting Test; TMT-B = Trail Making Test Form B; TS = Total Score; TTC = Time to Completion; VBRT = Visuospatial Block Recall Test; WLRT = Word List Recall Test.Note:
Reference: NOPRODMDD0003 CSR (TEFCOG01A.rtf)

Table 11: Test-Retest Reliability

Test	High value denotes better performance	r
SST NCC	Yes	0.769
SST % PE	Yes	0.665
DSFT Raw Score	Yes	0.637
DSBT Raw Score	Yes	0.715
WLRT-Rater LS	Yes	0.620
WLRT-Rater DRS	Yes	0.839
VBRT Num DCR	Yes	0.689
SDMaT TS	Yes	0.884
TMT-B TTC	No	0.469
BMT Num Errors	No	0.660

Abbreviation: BMT = Block Maze Test; CI = Confidence Interval; DCR = Designs Correctly Recalled; DRS = Delayed Recall Score; DSBT = Digit Span Backward Test; DSFT = Digit Span Forward Test; ICC = Intraclass Correlation Coefficient; LS = Learning Score; NCC = Number of Categories Completed; PE = Perseverative Errors; R=Revere.D; SD = standard deviation; SDMaT = Symbol Digit Matching Test; SST = Symbol Sorting Test; TMT-B = Trail Making Test Form B; TS = Total Score; TTC = Time to Completion; VBRT = Visuospatial Block Recall Test; WLRT = Word List Recall Test.Note:
r values obtained from NOPRODMDD0003 CSR (TEFTRR01.rtf)

7. REFERENCES

Carney CE, Buysse DJ, Ancoli-Israel S, et al. The Consensus sleep diary: Standardizing prospective sleep self-monitoring. Sleep 2012;35(2):287-302.

EuroQol Group. About EQ-5D. https://euroqol.org/eq-5d-instruments/ (Accessed 7 Nov 2023).

EuroQol Group. EQ-5D-5L User Guide: basic information on how to use the EQ-5D-5L instrument. Version 3 .0; Sep 2019. https://euroqol.org/publications/user-guides/ (Accessed 7 Nov 2023).

Guy W. CGI: Clinical Global Impressions. In: ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health and Human Services; printed 1976 (reprinted 1991):217-222.

Hunt SM, McKenna S. The QLDS: A scale for the measurement of quality of life in depression. Health Policy. 1992;22(3):307-319.

Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med. 1997;27(2):93-105.

Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia; 1985; 28:412–419.

McGahuey CA, Gelenberg AJ, Laukes CA, et al. The Arizona Sexual Experience Scale (ASEX): reliability and validity. J Sex Marital Ther. 2000;26(1):25-40.

Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382-389.

Nolen-Hoeksema S. The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. J Abnorm Psychol. 2000;109(3):504-511.

Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry. 2007;164:1035-1043.

Rickels K, Garcia-Espana R, Mandos LA, Case GW. Physician Withdrawal Checklist (PWC-20). J Clin Psychopharmacol. 2008;28(4):447-451.

Rubin, D. (1987), Multiple Imputation for Nonresponse in Surveys, Wiley, New York, XXIX+258 pp.

Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol. 1996 Jun;11 Suppl 3:89-95.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). [2009] Guidance for industry drug induced liver injury: premarketing clinical evaluation. Available: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM174090.pdf.

Yu L, Buysse DJ, Germain A, et al. Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. Behav Sleep Med. 2011;10(1):6-24.