UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Statistical Analysis Plan

Version: 1.0 Date: 16.07.2021

International Deprexis Trial in Multiple Sclerosis (IDEMS) – a multicenter randomized controlled trial

IDEMS

Study Protocol	Version 1.1 / 19.02.2017
Clinicaltrials.gov-No.	NCT02740361
Internal Protocol ID No.	Internal protocol number at the CTU-UMG
NCT-No.	NCT02740361
SAP Version	1.0 / 25.06.2021
Development Phase	Phase III
Coordinating Center	Charité – Universitätsmedizin Berlin NeuroCure Clinical Research Center (NCRC) Charitéplatz 1, 10117 Berlin
Coordinating Investigator	Prof. Dr. Stefan M. Gold Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin Charité – Universitätsmedizin Berlin Charitéplatz 1, 10117 Berlin

Approval of the Statistical Analysis Plan

International Deprexis Trial in Multiple Sclerosis (IDEMS) – a multicenter randomized controlled trial

NCT No.: NCT02740361

Protocol Version No: 1.0 / 16.07.2021

Prof. Dr. Stefan Gold

19.7.2021
Date

Signature

Biostatistician

Prof. Dr. Tim Friede

09 AUG 2021
Date

Signature

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AE	Adverse Event
ANCOVA	Analysis of Covariance
BDI-II	Beck Depression Inventory – II
BICAMS	Brief Cognitive Assessment in MS
BVMT-r	Brief Visuospatial Memory Test-revised
CFS	Chalder Fatigue Scale
CI	Confidence Interval
CVLT-II T1-5	California Verbal Learning Test-II
EDSS	Expanded Disability Status Scale
FAS	Full Analysis Set
FSMC	Fatigue Scale for Motor and Cognitive
ITT	Intention-To-Treat
LOCF	Last Observation Carried Forward
MADRS	Montgomery-Asberg Depression Rating Scale
MDD	Major Depressive Disorder
MI	Multiple Imputation
M.I.N.I.	Mini-International Neuropsychiatric Interview
mITT	modified Intention-To-Treat
MS	Multiple Sclerosis
MSIS-29	Multiple Sclerosis Impact Scale-29
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SBQ-R	Suicide Behaviors Questionnaire – Revised
SDMT	Symbol Digit Modalities Test
SS	Safety Set
SUSAR	Suspected Unexpected Serious Adverse Event
WHO	World Health Organization
WHO-QoL BREF	World Health Organization Quality of Life BREF
WLC	Waitlist Control Group
PDDS	Patient Determined Disease Steps

1 Introduction

This document has been written based on information contained in the trial protocol version 1.1, 02/2017. It describes the analysis after the first 12 weeks. The analysis after 12 months will be listed in a second Statistical Analysis Plan (SAP).

1.1 Background and Rationale

Depression is the most common comorbidity of Multiple Sclerosis (MS). However, depression remains underdiagnosed and there are no treatments with proven effectiveness from large (phase III) trials. In addition, symptoms of MS such as mobility issues, cognitive impairment and fatigue make it difficult for MS patients to travel to and attend regular psychotherapy. Effective treatment options are therefore needed. The internet-based program “Deprexis”, developed by GAIA group in Hamburg/Germany could facilitate access to treatment and has shown promising results in a small monocentre study in Germany for MS patients (1). Deprexis is an online tool and consists of 10 sequential modules plus an introduction and a summary module. Deprexis implements the technique of simulated dialogue by giving the user multiple choice options and tailoring the subsequent content to the patient’s responses. Thereby, the user’s responses determine the course of each module. The Deprexis program has been adapted from the generic version to improve suitability for use by patients with MS. This trial evaluates the effectiveness of MSDeprexis in an international multicentre trial. There are three treatment groups, one group uses only MSDeprexis for 12 weeks, the other group uses MSDeprexis with an added email support for 12 weeks and the third group is a waitlist control group (WLC).

1.2 Objective and endpoints

Table 1 Objectives and related endpoints

	Objective	**Endpoint**	**Measurement time points**
Primary	Assessment of **depressive symptoms** after intervention	Beck Depression Inventory – II (BDI-II)	V1 (baseline), I1 (after 4 weeks), I2 (after 8 weeks), V2 (after 12 weeks)
Secondary	Assessment of **quality of life** after intervention	World Health Organization Quality of Life BREF (WHO-QoL BREF) and the Multiple Sclerosis Impact Scale (MSIS-29) after 12 weeks	V1 (baseline), V2 (after 12 weeks)
	Assessment of **fatigue** after intervention	Fatigue Scale for Motor and Cognitive Functions (FSMC) and the Chalder Fatigue Scale (CFS) after 12 weeks	V1 (baseline), V2 (after 12 weeks)
	Assessment of the **percentage** of patient with a clinical diagnosis	M.I.N.I. structural clinical interview after 12 weeks	V1 (baseline), V2 (after 12 weeks)

	Objective	Endpoint	Measurement time points
	of current **major depressive disorder (MDD)** after intervention		
	Assessment of **severity of depression** after intervention	Clinician-rated Montgomery-Asberg Depression Rating Scale (MADRS) after 12 weeks	V1 (baseline), V2 (after 12 weeks)
Safety	Assessment of suicidal ideation	Suicide Behaviors Questionnaire – Revised (SBQ-R), predefined criterion for acute risk of suicide is response 3a or 3b on SBQ-R item 3 plus a score of 5 or 6 on SBQ-R item 4	V1 (baseline), V2 (after 12 weeks)

1.3 Primary objective and endpoint

To determine if MSDeprexis is effective for reducing **depressive symptoms** at the end of treatment (week 12), the Beck Depression Inventory – II (BDI-II)) in patients with MS is used and compared to the control group. Because the literature strongly suggests that therapist-guided internet-based interventions lead to better outcome MSDeprexis with an added email support (MSDeprexisPlus) is also compared to the control group after 12 weeks (MSDeprexis vs. WLC and MSDeprexisPlus vs. WLC).

1.4 Potential moderators

The level of disability measured by the Patient Determined Disease Steps (PDDS), the cognitive impairment measured by Symbol Digit Modalities Test (SDMT) and the level of fatigue measured by the FSMC will be investigated as potential moderators.

2 Study methods

2.1 Trial design

IDEMS is a three arm, randomized, controlled, international multicentre trial with five academic centres. Patients are randomized to one of the three treatment groups after informed consent. One group uses only MSDeprexis for 12 weeks, the other group uses MSDeprexis with an added email support for 12 weeks and the third group is a waitlist control group.

The potential of a booster session to enhance maintenance is analysed in patients who were randomized to either the MSDeprexis or the MSDeprexisPlus group. These patients will be re-randomized to receive additional booster session (see Figure 1). This analysis is not part of this SAP.

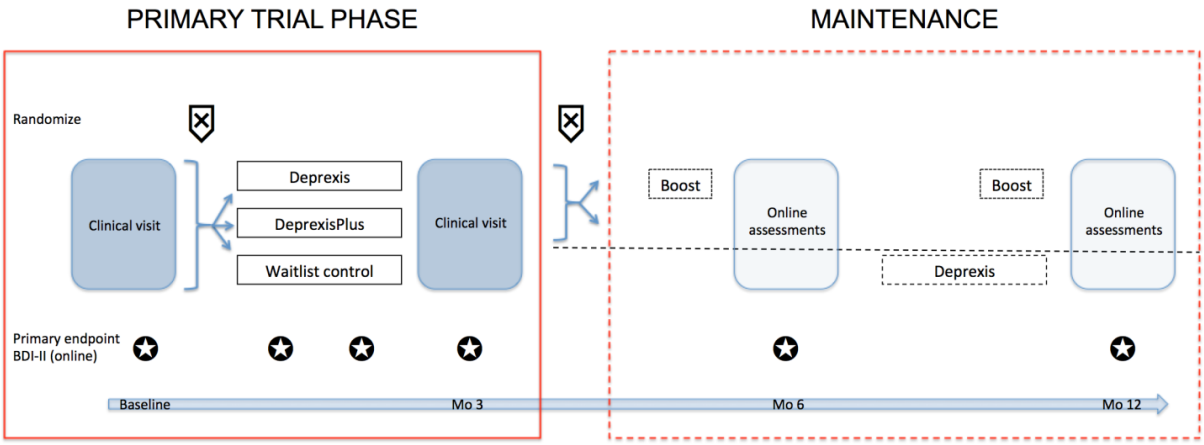

Figure 1: Trial Design Schematic and Treatment Plan.

2.2 Randomization

Patients are randomized 1:1:1 to one of the three trial arms. There is no blocking and no stratification. To ensure concealed allocation, eligibility is determined and all baseline assessments completed before randomization in compliance with CONSORT guidelines. The clinicians and raters who will be conducting clinical assessments (structured interviews) will be blind to treatment assignment (single blind RCT).

2.3 Sample Size

A sample size of 100 patients per intervention group gives a conjunctive power (probability of rejecting both null hypotheses comparing MSDeprexis and MSDeprexisPlus to waitlist control) of 90% for a Dunnett test at the usual one-sided significance level of 2.5% assuming standardized mean differences of 0.5 for MSDeprexis vs. WLC and 0.8 for MSDeprexisPlus vs. WLC in the primary outcome change in BDI from baseline to week 12. Adjusting for 20% dropout we aim to recruit 125 patients per group resulting in a total sample size of 375 patients. The power was simulated with 10,000 replications using EAST 6.3.

2.4 Framework

The superiority of MSDeprexis compared to a waitlist control group is analysed. Specifically, the use of MSDeprexis only as well as the use of MSDeprexis with an added email support are compared to the control group.

2.5 Statistical Interim Analyses and Stopping Guidance

No interim analyses or guidelines for stopping the trial early are planned. There are no predefined stopping rules.

2.6 Timing of the Final Analysis

All outcomes are analysed collectively after the last patient has completed the visit after 12 weeks. In addition, to explore the long-term stability of therapeutic effects and the potential of a booster session to enhance maintenance another evaluation takes place after 12 months (not part of this SAP).

2.7 Timing of Outcome Assessments

Outcomes are measured at week 0 (baseline) and week 12 when visiting the clinic. In addition, BDI-II outcomes are measured at week 4 and 8 in online assessments. Follow-up visits for all patients who want to use MSDeprexis after 6 and 12 month are conducted online.

3 Statistical Principles

3.1 Confidence intervals and p-values

If not specified otherwise, tests will be performed two-sided with a significance level of 5% and 95% confidence intervals (CI) will be provided for parameter estimates. For the primary analysis, MSDeprexis vs. WLC and MSDeprexisPlus vs. WLC will be tested by a Dunnett test controlling the familywise type I error rate at the level of 2.5% (one-sided).

3.2 Adherence and protocol deviations

The number of days with activity in the MSDeprexis programm by each patient is assessed as a measure of treatment adherence.

3.3 Analysis populations

Full analysis set (FAS)

The FAS is based on the intention-to-treat (ITT) principles. This means that all randomized patients with at least one post-baseline assessment will be included in the analysis. A modified ITT (mITT) population will be analysed in a sensitivity analysis. It includes those patients who have registered in the program MSDeprexis.

Safety Set (SS)

All subjects who have registered in the program MSDeprexis will be included. Subjects will be analysed according to the treatment they received.

4 Trial population

4.1 Screening data

Screening takes place within the participating centres. Available data on patients assessed for eligibility will be included within a CONSORT flow diagram.

4.2 Eligibility

Patients need to be at least 18 years old with a neurologist-confirmed diagnosis of MS. They have self-reported depressive symptoms (BDI-Fastscreen ≥ 4) and are able to travel to the outpatient centre for two clinical assessments. Acute risk for suicide at baseline will lead to exclusion from the trial. For complete inclusion/exclusion criteria, see section 4 of the latest version of the study protocol.

4.3 Recruitment

A CONSORT flow diagram will be calculated to show patient disposition including sample sizes for recruitment, randomization and analyses.

4.4 Withdrawal/follow-up

Withdrawal rates will be calculated and reasons given within a CONSORT flow diagram.

4.5 Baseline patient characteristics

Baseline characteristics and demographic data will be summarized descriptively stratified by treatment group:

- Age [years]
- Sex [female; male]
- Marital status [married of domestic partnership; separated or divorced; single, never married; widowed]
- Education [high school graduate, diploma or equivalent; associate degree, bachelor's degree; master's degree; doctoral degree]
- Employment status [unemployed or retired; homemaker; student, parttime; full time; other]
- Time since diagnosis
- Diagnosis subtype [primary progressive MS; secondary progressive MS; relapsing-remitting MS; unclear]
- Neurological status [Expanded Disability Status Scale (EDSS)]
- Level of disability [PDDS]
- Neurological function [BICAMS]

- Disease-modifying therapy.

5 Analysis

5.1 Outcome definitions

Depressive symptoms at baseline and at week 4, 8 and 12 (end of treatment) are measured with the BDI-II. BDI-II is a 21-question self-reported inventory for measuring the severity of depression. Each question has a set of at least four possible responses, ranging in intensity. To calculate the score a value of 0 to 3 is assigned for each answer. A higher total score indicates more severe depressive symptoms. The interpretation of these scores is:

- 0-13: minimal depression
- 14-19: mild depression
- 20-28: moderate depression
- 29-63: severe depression (2)

Quality of life is measured at baseline and at week 12 with the WHO-QoL BREF and with the MSIS-29. WHO-QoL BREF consists of 26 items, which are assigned to the four domains: physical health, psychological health, social relationships, and environment. Answers are on a 5-points Likert scale ranging from 1 (not at all) to 5 (completely). Items are scored from 1 to 5 and a score (0-100) is calculated using an algorithm (see Appendix 10 of (3)).
The Multiple Sclerosis Scale-29 includes two scales: physical impact (20 items) and psychological impact (nine items). All items have five response options from 1 (not at all) to 5 (extremely). A score (0-100) is calculated using the following algorithm: Physical scale items (1-20): sum, subtract 20, divide by 80, and multiply by 100, psychological scale items (21-29): sum, subtract nine, divide by 36, and multiply by 100 (4).

Fatigue is measured at baseline and at week 12 with the FSMC and the Chalder Fatigue Scale. FSMC is a 20-item scale with response option from 1 (Does not apply at all) to 5 (Applies completely). There are ten questions about the cognitive and ten questions about the motor fatigue, the total score can reach 100 points (extreme fatigue) (5).
The Chalder Fatigue Scale is an 11-item questionnaire with a 4-points Likert scale ranging from 0 (Less than usual) to 3 (Much more than usual). The sum of these items is the score, so it can range from 0 to 33 (6).

In addition, depressive symptoms are measured with the Mini-International Neuropsychiatric Interview (M.I.N.I.) and the MADRS. M.I.N.I. is a short structured diagnostic interview for the major Axis I psychiatric disorders in DSM-IV and ICD-10 (7).
MADRS is a ten-item questionnaire with a 7-points Likert scale from 0 to 6. The score is the sum of all items and ranges from 0 to 60 (8).

For safety analysis the SBQ-R is used to control suicidal ideation. The SBQ-R has 4 questions, where each question has an individual scale, and each response corresponds to a certain point value. The total score ranges from 3 to 18. The higher the score, the higher the likelihood of suicidal behavior (9).

To determine the neurological function the Symbol Digit Modalities Test (SDMT) is used. This test is part of the Brief Cognitive Assessment in MS (BICAMS). The SDMT consists of a row with nine symbols, where each symbol corresponds to a certain value. Below there are seven rows with 15 symbols each and the patient should assign the correct numerical value to as many symbols as possible within 90 seconds. The score is the number of correct assignments. There are two additional tests: The California Verbal Learning Test-II (CVLT-II T1-5). There is a 16-item list with words from four different topics, arranged randomly. The list is read aloud five times in the same order to the patient. Patients are required to recall as many items as possible, in any

order, after each reading of the list. The other one is the Brief Visuospatial Memory Test-revised (BVMT-r). In this test 6 abstract figures are presented three times. In each of the three learning trials, the patient views the same array for 10 seconds. Then the array is removed and the patient is required to draw the stimulus array from memory, with the correct shapes in the correct position. Each design receives from 0 to 2 points representing accuracy and location, so the score ranges from 0 to 12 for each trial (10, 11).

To determine the neurological impairment the Patient Determined Disease Steps (PDDS) is used. There are 9 categories from 0 (no limitation of activities) to 8 (bedridden) describing the patients situation with a main focus on the ability to walk (12).

5.2 Analysis methods

Primary endpoint

The primary outcome change in BDI from baseline to week 12 will be analysed by means of linear mixed effects models for repeated measures adjusted for baseline measurements with fixed effects for intervention, region (US, GER), time and baseline BDI score, and random subject effects for individual patients including all patients with at least one post-baseline measurement (13). Least squares means will be reported for the intervention groups with 95% CI as well as the difference between the least squares group means with 95% CI. MSDeprexis vs. WLC and MSDeprexisPlus vs. WLC will be tested by a Dunnett test controlling the familywise type I error rate at the level of 2.5% (one-sided). The primary hypothesis will be tested in a confirmatory manner:

$H_0^{(1)}: \mu_{WLC} \leq \mu_{Dep}$ vs. $H_1^{(1)}: \mu_{WLC} > \mu_{Dep}$ and $H_0^{(2)}: \mu_{WLC} \leq \mu_{DepPlus}$ vs. $H_1^{(2)}: \mu_{WLC} > \mu_{DepPlus}$,

with μ_{WLC}, μ_{Dep} and $\mu_{DepPlus}$ being the means of the change in BDI from baseline to week 12.

In a secondary step, the added value of therapist email support MSDeprexis vs. MSDeprexisPlus at a two-sided level of 5%, if efficacy of MSDeprexis and MSDeprexisPlus for reducing depressive symptoms in MS could be demonstrated, will be determined.

SAS analysis:

```
proc mixed data =_data_t;
class group time region;
model score = group time group*time baseline baseline*time region / s ddfm=kr;
repeated time / subject=id type=un;
lsmeans group / adjust = dunnett;
```

In a sensitivity analyses last observation carried forward (LOCF) and multiple imputations (MI) will be used to deal with missing values for the BDI-II score (missing visits) and for each method an analysis of covariance (ANCOVA) for the BDI-II score after 12 weeks is carried out with BDI-II score at baseline as covariate.

In an exploratory analysis trends in the utilization time of MSDeprexis [number of days with activity within the program] in both intervention groups will be investigated.

Secondary endpoints

The analyses of secondary endpoints QoL (WHO-QoL and MSIS-29), fatigue (FSMC and CFS) and the severity of depression (MADRS) will follow the similar approach as the analyses described for the primary endpoint. The psychological score of the WHO-QoL and the FSMC are defined as the key secondary endpoints.

The number of patient with a clinical diagnosis of current major depressive disorder (M.I.N.I.) will be analysed using a logistic regression model with the variables treatment group and baseline score of BDI-II.

Potential moderators

The potential moderators (SDMT, PDDS and FSMC) will be investigated in a supporting exploratory analyses by including these and their interaction with treatment in the linear mixed effects models described above.

Safety analyses

The analysis of suicidal ideation (measured by SBQ-R) is done analogously to the primary endpoint.

5.3 Missing data

Five patients have no online baseline assessment (technical problems). For the primary analysis these missing baseline BDI-II scores will be replaced by a regression imputation with the MADRS scores which were collected paper based. Although the mixed model described above is robust to a certain extent to missing data, sensitivity analyses will be performed as supporting analyses (see section 5.2).

5.4 Harms

Adverse events will be summarized as frequencies and percentages by intervention group. New occurrence of suicidal ideation or intent as well as worsening of depressive symptoms above the clinical threshold are considered as potential adverse events. Every AE fulfilling one of the following criteria is a SAE:

- Suicidal ideation or intent (as measured by a score of 3 on BDI-II item 9 at any assessment, during the clinical interview at V2 or spontaneous report in a web message to the therapist or by contacting the study site)
- Hospitalization due to psychiatric disorder classified according to ICD10 or DSM5
- Lethal or life-threatening (incl. suicide or suicide attempt).

However, the intervention is generally considered low risk as Deprexis has been used in numerous clinical trials without evidence for adverse events and is categorized as a “low risk” medical device according to its German CE®-certification.

5.5 Statistical software

Statistical programming will be done using R version 4.0.0 or higher and SAS version 9.4 or higher.

5.6 References

1. Fischer, A., Schröder, J., Vettorazzi, E., Wolf, O. T., Pöttgen, J., Lau, S., ... & Gold, S. M. (2015). An online programme to reduce depression in patients with multiple sclerosis: a randomised controlled trial. *The Lancet Psychiatry*, *2*(3), 217-223.
2. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.
3. World Health Organization (1998). Programme on metal health – WHOQOL User Manual.
4. Widener, G. L., & Allen, D. D. (2014). Measurement characteristics and clinical utility of the 29-item Multiple Sclerosis Impact Scale. *Archives of physical medicine and rehabilitation*, *95*(3), 593-594.
5. Penner, I. K., Raselli, C., Stöcklin, M., Opwis, K., Kappos, L., & Calabrese, P. (2009). The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. *Multiple Sclerosis Journal*, *15*(12), 1509-1517.
6. Jackson, C. (2015). The Chalder fatigue scale (CFQ 11). *Occupational Medicine*, *65*(1), 86-86.
7. Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.
8. Montgomery, SA (1979). Åsberg M. *A new depression scale designed to be sensitive to change. Br J Psychiatry*, *134*, 382-389.
9. Osman, A., Bagge, C. L., Gutierrez, P. M., Konick, L. C., Kopper, B. A., & Barrios, F. X. (2001). The Suicidal Behaviors Questionnaire-Revised (SBQ-R): validation with clinical and nonclinical samples. *Assessment*, *8*(4), 443-454.
10. Langdon, D. W., Amato, M. P., Boringa, J., Brochet, B., Foley, F., Fredrikson, S. & Reder, A. T. (2012). Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis Journal*, *18*(6), 891-898.
11. Langdon, D., Amato, M., Boringa, J., Brochet, B., Foley, F., Fredrikson, S. & Reder, A. (2011). The Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS): first consensus steps towards a brief universal cognitive assessment for MS. In *Neurology* (Vol. 76, No. 9, pp. A479-A479). 530 WALNUT ST, PHILADELPHIA, PA 19106-3621 USA: LIPPINCOTT WILLIAMS & WILKINS.
12. Hohol, M. J., Orav, E. J., & Weiner, H. L. (1995). Disease steps in multiple sclerosis: a simple approach to evaluate disease progression. *Neurology*, *45*(2), 251-255.
13. Mallinckrodt CH, Clark SW, Carroll RJ, Molenbergh G (2003). Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. Journal of biopharmaceutical statistics. 13:179-190.
14. Henderson R, Diggle P, Dobson A (2000). Joint modelling of longitudinal measurements and event time data. Biostatistics. 1:465-480.