Statistical Analysis Plan (SAP)

Full Title of Study	Combined N-of-1 Trials to Assess Open-Label-Placebo Treatment for Antidepressant Discontinuation Symptoms
Short Title	FAB-Study
Principal Investigator	Yvonne Nestoriuc, PhD Professor for Clinical Psychology and Psychotherapy Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg, Clinical Psychology and Psychotherapy, Hamburg, Germany & University Medical Center Hamburg-Eppendorf, Institute of Systems Neuroscience, Hamburg, Germany
Protocol Identification number	0721-20
ClinicalTrials.gov identifier	NCT05051995
Author	Responsible statistician: Stefan Konigorski, PhD Digital Health – Machine Learning Group, Hasso Plattner Institute for Digital Engineering, Potsdam, Germany & Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA & Department of Statistics, Harvard University, 150 Western Ave, Boston, MA 02134, USA Responsible employee: Amke Müller, MSc Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg, Clinical Psychology and Psychotherapy, Hamburg, Germany & University Medical Center Hamburg-Eppendorf, Institute of Systems Neuroscience, Hamburg, Germany
Version	01
Sponsor	Deutsche Forschungsgemeinschaft (DFG)

Signature Page

Principal Investigator _____

Author _____

_____

Document History:

Version No.	Date	Chapter / Attachment	Reason for Review / Modification
First draft	21.12.2023	n.a.	n.a.

Content

1. Introduction

The aim of this project is to investigate the efficacy of open-label-placebo (OLP) treatment in reducing antidepressant discontinuation symptoms in remitted depressive disorder using a series of N-of-1 trials. This Statistical Analysis Plan (SAP) will provide more detailed description about the endpoints and corresponding analyses.

This study is part of a collaborative research center (CRC 289 Treatment Expectation: www.treatment-expectation.de/en/).

1.1. Background

Discontinuation of antidepressant medication is associated with a broad range of physical and psychological symptoms, with an average incidence of 56% in fully remitted cases (Davies and Read 2019). Clinical trials indicate that the therapeutic effect of antidepressant medication over and above placebo is relatively small, in other words, placebo effects show a similar benefit as active antidepressants in antidepressant trials (Cipriani et al. 2018). Open-label-placebo (OLP) treatment appears to be effective for symptoms common during discontinuation of antidepressant medication such as pain, flu and fatigue (Wernsdorff et al. 2021). This study aims to investigate the efficacy of OLP in reducing antidepressant discontinuation symptoms through a series of N-of-1 trials. N-of-1 trials offer a pragmatic methodology for the proposed study as they require a smaller sample size compared to randomized controlled trials. Therefore, they are particularly suited for populations that are challenging to recruit (i.e., patients fulfilling the guideline criteria for discontinuation of their antidepressant medication) (Guyatt et al. 2000). They allow to obtain estimates of individual-level treatment effects and aggregating a series of N-of-1 trials allow for estimates population-level treatment effects (Zucker et al. 1997).

1.2. Overall Study Aims

Aim 1: To obtain individual-level and population-level effect estimates of the average efficacy of OLP relative to no treatment in reducing antidepressant discontinuation symptoms (primary outcome), and in improving dysfunctional treatment expectation, depressed mood, and anhedonia (secondary outcomes) based on a series of N-of-1 trials.

Aim 2: To explore factors that modify the efficacy of OLP, i.e., age, female sex, prior negative experiences with discontinuation, discontinuation symptom load over the discontinuation period and higher maintenance dose and longer duration of antidepressant use.

1.3. Study Endpoints

1.3.1. Primary Endpoint

Discontinuation symptoms during the experimental phase (N-of-1 trials) are assessed twice daily for eight weeks (112 assessments) via a smartphone application (‘StudyU’) using the

Generic Rating Scale for Treatment Effects ($GEEE_{ACT}$) adjusted for the ambulatory assessment context (Konigorski et al. 2022; Rief et al. 2021). Discontinuation symptoms are rated on a 11-point numeric rating scale (NRS) ranging from 0-10. Higher scores indicate more severe discontinuation symptoms.

1.3.2. Secondary Endpoints

Treatment expectations during the experimental phase (N-of-1 trials) are assessed twice daily for eight weeks via the 'StudyU' application using an adjusted version of the Generic Rating Scale for Treatment Expectations ($GEEE_{EXP}$) (Konigorski et al. 2022; Rief et al. 2021). Treatment expectations are rated on a 11-point NRS, with higher scores indicating expectation of more severe discontinuation symptoms until the next assessment. Depressed mood and anhedonia during the experimental phase are measured via the 'StudyU' application using the Patient-Healthcare-Questionnaire-2 (PHQ-2) consisting of 2 items adjusted to the ambulatory context (Kroenke et al. 2003). These are rated on a 4-point Likert scale, ranging from 0-3, with higher scores suggestive of a higher degree of depressive symptomatology.

1.3.3. Safety Endpoints

Major safety concerns for participants discontinuing their antidepressant medication include discontinuation symptoms, depressive symptoms, and adverse events (AEs). All safety endpoints are monitored throughout the study at (bi-)weekly study visits. Discontinuation symptoms are assessed with the Generic Rating Scale for Treatment Effects ($GEEE_{ACT}$) (Rief et al. 2021). Depressive symptoms will be monitored by self- and expert-rated depressive symptoms with the Beck Depression Inventory (BDI-II) and the Montgomery-Åsberg Depression Scale (MADRS) (Hautzinger et al. 2006; Davidson et al. 1986). AEs will be examined by a single safety question and classified according to the 'Common Terminology Criteria for Adverse Events' (CTCAE) (National Institutes of Health, National Cancer Institute 2009). In addition, the assessor evaluates the clinical impression of the participant.

2. Study Methods

2.1. Trial Design

This trial is a series of randomized, single-blinded, multicenter N-of-1 trials that compares OLP to no treatment in reducing antidepressant discontinuation symptoms. Treatment periods consist of 2-week periods of OLP and no treatment in an alternating order (i.e., ABAB/BABA), where 'A' represents OLP and 'B' no treatment. Participants will be randomized to either OLP or no treatment as their starting intervention and alternate between these two conditions during the eight-week experimental phase (N-of-1 trials). Results of the individual N-of-1 trials will be combined to estimate the average population-level efficacy of OLP compared to no treatment.

2.2. Randomization and Blinding

Patients will be randomized with a 1:1 ratio to the two conditions differing in treatment order (ABAB/BABA). The allocation sequence is based on computer-generated random numbers in R-Studio with the package *blockrand* using a block-randomization. Patients and study staff responsible for the randomization procedure will be unblinded to treatment allocation. Unblinded study staff are not involved in further assessments, except for the instructions of the treatment conditions. All other researchers involved in the assessments of the patients will be blinded to treatment allocation and patients are asked not to inform the assessor about the current condition.

2.3. Sample Size

Sample size are based on the main analysis concerning the aggregation of the N-of-1 trials to assess the average treatment efficacy of OLP compared to no treatment in reducing discontinuation symptoms assessed on a 11-point numeric rating scale $GEEE_{ACT}$ (Rief et al. 2021). Sample size calculation was performed using the *Shiny-App* (https://jiabeiyang.shinyapps.io/SampleSizeNof1/) which implements a linear mixed model specifically designed for sample size calculations in a series of N-of-1 trials (Yang et al. 2021). We assumed a fixed intercept and random slope model for alternating sequences (ABAB/BABA). We used the daily $GEEE_{ACT}$ scores for a fixed number of patients (N = 20) and a fixed number of measurements per patient (112 measurements) with a standardized homogeneous residual error of 1.5, variance of random slope of 0.75, and an AR-1 structure for repeated measures with a correlation of 0.7 (conservative estimation). The results of the sample size calculation suggest that a sample size of 18 participants yields a power of 93% for identifying a mean difference between OLP treatment and no treatment when the mean difference is 0.8 points (on the $GEEE_{ACT}$ 11-point numeric rating scale) which we consider as minimally clinically important difference (at a significance level of 5%). Considering possible dropout, we will recruit 20 patients into the study. This study design can be expected to yield naïve (i.e., non-pooled) estimates of the individual treatment effects with a standard error of about 0.3, hence allowing also meaningful inference on the individual level.

2.4. Eligibility

Inclusion criteria

- Adult patients (≥18) meeting lifetime criteria for a major depressive disorder (MD) diagnosis, according to the Structured Clinical Interview for DSM-V (SCID-V-CV) (First et al. 2016);
- Current maintenance treatment with citalopram (20-40mg), duloxetine (60-100mg), escitalopram (10-20mg), paroxetine (20-40mg), sertraline (50-150mg), venlafaxine (75-150mg) or mirtazapine (30-45mg) with a constant dosage for 4 weeks;

- Discontinuation wish by participant supported by prescribing physician;
- Fulfilment of the S3 German national guideline recommendations to discontinue antidepressant medication: a) response to antidepressant medication; b) symptom remission ≥4 months (first episode) or ≥2 years (≥2 episodes with significant functional impairment);
- Informed consent;
- Additional criteria for N-of-1 trial: At least moderate discontinuation symptoms after antidepressant discontinuation assessed by the Generic Rating Scale of Treatment Effects ($GEEE_{ACT}$ score ≥4 during past week).

Exclusion criteria

- Current moderate or severe psychopathological symptoms or psychosocial impairments;
- Acute or chronic somatic illness which might interfere with depressive disorder, antidepressant or proposed study;
- Acute suicidality, psychotic symptoms, substance abuse, or addiction, current mania, or hypomania confirmed by SCID-V-CV or other psychopathology which might interfere with depressive disorder, antidepressant or proposed study;
- Any history of bipolar disorder or psychosis confirmed by SCID-V-CV;
- Severe stressful life events within 6 months prior to study participation;
- Current pregnancy;
- Insufficient German language proficiency.

2.5. Timing of Outcome Assessments

Table 1: Schedule of enrollment, intervention, and assessments according to SPIRIT-PRO

	STUDY PERIOD						
	Phase 1: Screening & Eligibility		**Phase 2: Discontinuation**		**Phase 3: N-of-1 trials**		**Phase 4: Follow-up**
Timepoint	**S1**	**S2**	**T0**	**T1-T5**	**T6-T8**	**T9**	**FU1**
Week	**-2**	**-1**	**0**	**1/2/3/4/5**	**7/9/11**	**13**	**26**
Enrolment:							
Eligibility screen	✓	✓					
Informed consent		✓					
Initial consultation prescribing physician		✓					
Allocation (N-of-1 trials)				✓ (T5)			
Interventions:							

Run-In (1 week)		◆—◆				
Discontinuation phase (4 weeks)			◆—◆			
N-of-1 trials (8 weeks)				◆—◆		
Assessments:						
Primary outcome **('StudyU')**						
Discontinuation symptoms ($GEEE_{ACT}$)				◆—◆		
Secondary outcomes **('StudyU')**						
Treatment expectations ($GEEE_{EXP}$)				◆—◆		
Depressed mood and anhedonia (PHQ-2)				◆—◆		
Effect modifiers **('LimeSurvey' and clinical interviews)**						
Gender		✓				
Maintenance dose of antidepressant medication	✓					
Duration of antidepressant use	✓					
Prior negative experiences with discontinuation of antidepressant medication ($GEEE_{PRE}$)*		✓				
Discontinuation symptom load over the discontinuation period (DESS)		✓	✓			
Other pre-specified outcomes **('LimeSurvey' and clinical interviews)**						
Discontinuation symptoms (DESS)		✓	✓	✓	✓	✓
Current treatment effects ($GEEE_{ACT}$)			✓	✓	✓	✓
Expert-rated depressive symptoms (MADRS+)*	✓	✓	✓	✓	✓	✓

Self-reported depressive symptoms (BDI-II)*	✓	✓	✓	✓	✓	✓
Recurrence			✓	✓	✓	✓
Well-being (SWEMWBS)		✓	✓	✓	✓	✓
Adherence (single item)		✓	✓	✓	✓	✓
Adverse events (single safety items)		✓	✓	✓	✓	✓
Treatment expectations (TEX-Q, $GEEE_{EXP}$)*		✓	✓	✓	✓	✓ ($GEEE_{EXP}$)
Psychopathology (SCID-5-CV)*	✓					
Demographic* & medical characteristics	✓	✓				
Anxiety vs. Depression (STADI Trait/State)*		✓				
Psychophysiological stress (PSS-10)*		✓				
Prior treatment experience ($DESS_{PAST}$)		✓				
Side-effects of antidepressant medication (GASE)*		✓			✓	
Antidepressant medication blood serum level			✓ (T1)		✓	

Note. S = Screening; *T* = Treatment; *FU* = Follow-Up; *DESS* = Discontinuation Emergent Signs and Symptoms Scale; $GEEE_{ACT}$ = Generic Rating Scale for Treatment Effects; *BDI-II* = Beck-Depressions-Inventory-II; *MADRS* = Montgomery-Åsberg Depression Rating Scale; *SWEMWBS* = Short Warwick-Edinburgh Mental Wellbeing Scale; *SCID-V-CV* = Structured Clinical Interview for DSM-V - Clinician Version; *STADI* = State-Trait Anxiety-Depression Inventory; *TEX-Q* = Treatment Expectation Questionnaire; $GEEE_{EXP}$ = Generic Rating Scale for Treatment Expectations; $GEEE_{PRE}$ = Generic Rating Scale for Previous Treatment Experiences; *GASE* = Generic Assessment of Side-Effects; *PHQ-2* = Patient-Health-Questionnaire-2; *PSS-10* = Perceived Stress Scale-10. *Part of standardized psychometric test battery.

3. Statistical Analysis

3.1. Trial Profile

A CENT (CONSORT extension (Shamseer et al. 2016)) flow diagram will display the flow of patients through the study. We will report the number of patients who (1) were assessed for

eligibility at screening*, (2) were included in the study*, (3) met the inclusion criteria for the N-of-1 trial*, (4) allocated to each randomization sequence*, (5) received the allocated treatment*, (6) included in the primary analysis*.

*Reasons for exclusion will be provided

3.2. Baseline Characteristics of Patients

Baseline characteristics of patients will be described including demographic parameters and medical characteristics. Categorical variables will be summarized by frequencies and percentages. Continuous variables will be reported by mean, standard deviation, minimum, maximum, or median and percentiles.

3.3. Analysis Principles for Final Analyses

The primary analyses will provide individual-level and population-level estimates of the average efficacy of OLP treatment in reducing discontinuation symptoms based on the series of N-of-1 trials. This analysis will be based on the intention-to-treat population.

3.3.1. Analyses of Individual N-of-1 Trials

At the end of the individual N-of-1 trials, statistical analyses will be applied to compare the outcomes of OLP vs no treatment. Bayesian models will be utilized to compare the average ratings of discontinuation symptoms (primary outcome) during the periods of OLP treatment periods to those with no treatment. These obtained estimates of individual-level effects can be viewed as naïve estimates as they do not incorporate information from the other participants. We will include a first-order autoregressive (AR1) error structure to acknowledge that measurements at adjacent times show a higher correlation than measurements far apart in time, given the longitudinal nature of our data. Non-informative priors will be selected for all parameters, as no prior information was available before the study that allowed a reasonable effect estimate of OLP treatment in reducing discontinuation symptoms. Separate analyses will be conducted for each individual to compare the average response to treatment. Based on this model, we will obtain estimates of the posterior distribution of the average OLP treatment effect in reducing discontinuation symptoms on an individual level. Secondary outcomes such as dysfunctional treatment expectations, depressed mood, and anhedonia will be analyzed similarly.

3.3.2. Analyses of the Aggregated N-of-1 Trials

For the aggregation of the N-of-1 trials, a Bayesian multi-level model will be applied to assess the efficacy of OLP relative to no treatment in reducing discontinuation symptoms (primary outcome) at population-level. We will utilize multi-level mixed models to estimate the posterior distribution of the population-level average treatment effect and the within- and between-patient variance. Additionally, we will obtain a pooled estimate of the individual-level patient's

treatment effect. For the aggregated analysis, noninformative priors will be selected for all parameters, as there was no prior information available before the study that allowed a reasonable effect estimate of OLP treatment in reducing discontinuation symptoms. The dependence of responses over time will be modeled with an AR1 error structure. Similarly, to the individual-level model described above, we will obtain estimates of the posterior distribution of the average OLP treatment effect on the population-level in reducing discontinuation symptoms. Extensions to the above model will be applied to account for time trends. Further, we will assess between-patient covariates such as age, female gender, prior negative pre-experiences with discontinuation, intensity of discontinuation symptoms over the course, higher maintenance dose, and duration of antidepressant use to determine if the given outcome is related to such variables. From the aggregated Bayesian analysis, we derive the posterior distribution of the mean difference between the outcomes of OLP and no treatment. Secondary outcomes will be modelled in the same manner. Statistical analyses will be performed in JAGS running from R, using the Markov Chain Monte Carlo (MCMC) method to obtain empirical samples from the joint posterior distribution of the parameters. Secondary outcomes such as dysfunctional treatment expectations, depressed mood, and anhedonia will be modelled in the same manner.

3.4. Protocol deviations

Treatment adherence to the intake of the OLP pills during the experimental phase (N-of-1 trials) is assessed at study visits by a single item inquiring how many times the placebo pill was taken since the last study visit.

Major protocol deviations include:

- Withdrawal of informed consent
- Pregnancy
- Insufficient treatment adherence regarding requirements for the study participation (intake of at least 80% of placebo pills during OLP treatment phase) (Nestoriuc et al. 2016)
- Restarting of antidepressant use during experimental phase.

3.5. Missing Data

We will disregard missing values in the analyses and perform complete-case analyses, hence we are assuming that missing values are missing completely at random. In case of major protocol deviations before the end of the N-of-1 trial, we will analyze all available data.

3.6. Analysis Populations

3.6.1. Intention-To-Treat (ITT)

The primary analyses are based on the ITT population, i.e., all randomized patients are included, regardless of whether they completed the whole study period and received the allocated intervention. For patients that dropped out of the study or that restarted the use of antidepressant medication during the N-of-1 trial, their recorded data until that timepoint will be included in the analyses.

3.6.2. Per Protocol (PP)

The sensitivity analysis is based on the per protocol population, i.e., all randomized patients, who have no major protocol violations and completed the whole study (i.e., compliance to ambulatory assessment >50% in the eight-week N-of-1 trial).

3.7. Safety and Adverse Events (AE)

Treatment will be terminated in case of patient's withdrawal of the informed consent, pregnancy and objections by study staff based on pre-defined health risks. These include treatment-related serious AEs (SAEs), recurrence of major depression, acute suicidality, and severe discontinuation symptoms. Safety-relevant outcomes will be evaluated at every study visit by trained study staff.

A data safety monitoring board (DSMB) will advise and review the proposed study. Study-related AEs, drop-outs due to AEs, medical events and mortalities will be reported immediately (within 48 hours). The DSMB will receive information about safety-relevant outcomes regularly.

4. Statistical Software

- JAGS version 4.3.0 or newer
- R version 4.2.2 or newer

5. Differences to trial protocol

6. References

Cipriani, A.; Furukawa, T. A.; Salanti, G.; Chaimani, A.; Atkinson, L. Z.; Ogawa, Y. et al. (2018): Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. In *Focus (Am Psychiatr Publ)* 16 (4), pp. 420–429. DOI: 10.1176/appi.focus.16407.

Davidson, J.; Turnbull, C. D.; Strickland, R.; Miller, R.; Graves, K. (1986): The Montgomery-Åsberg Depression Scale: reliability and validity. In *Acta Psychiatrica Scandinavica* 73 (5), pp. 544–548.

Davies, James; Read, John (2019): A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: Are guidelines evidence-based? In *Addictive Behaviors* 97, pp. 111–121. DOI: 10.1016/j.addbeh.2018.08.027.

First, Michael B.; Williams, Janet B. W.; Karg, Rhonda S.; Spitzer, Robert L. (2016): Structured Clinical Interview for DSM-5 Disorders-Clinician Version (SCID-5-CV): American Psychiatric Association.

Guyatt, G. H.; Haynes, R. B.; Jaeschke, R. Z.; Cook, D. J.; Green, L.; Naylor, C. D. et al. (2000): Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. In *JAMA* 284 (10), pp. 1290–1296.

Hautzinger, M.; Keller, F.; Kühner, C. (2006): BDI-II Beck Depressions-Inventar Revision. 2nd ed. Frankfurt am Main, Germany: Harcourt Test Services.

Konigorski, S.; Wernicke, S.; Slosarek, T.; Zenner, A. M.; Strelow, N.; Ruether, D. F. et al. (2022): StudyU: A platform for designing and conducting innovative digital N-of-1 trials. In *Journal of Medical Internet Research* 24 (7), e35884.

Kroenke, Kurt; Spitzer, Robert L.; Williams, Janet B. W. (2003): The Patient Health Questionnaire-2: validity of a two-item depression screener. In *Medical care* 41 (11), pp. 1284–1292. DOI: 10.1097/01.MLR.0000093487.78664.3C.

National Institutes of Health, National Cancer Institute (2009): US Department of Health and Human Services. Common terminology criteria for adverse events (CTCAE) version 4.0.

Nestoriuc, Y.; Blanckenburg, P. von; Schuricht, F.; Barsky, A. J.; Hadji, P.; Albert, U-S; Rief, W. (2016): Is it best to expect the worst? Influence of patients' side-effect expectations on endocrine treatment outcome in a 2-year prospective clinical cohort study. In *Annals of oncology : official journal of the European Society for Medical Oncology* 27 (10), pp. 1909–1915. DOI: 10.1093/annonc/mdw266.

Rief, W.; Nestoriuc, A. Y.; Mueller, Erik M.; Hermann, C.; Schmidt, Katharina; Bingel, U. (2021): Generic rating scale for previous treatment experiences, treatment expectations, and treatment effects (GEEE). In *PsychArchives*. DOI: 10.23668/PSYCHARCHIVES.4717.

Shamseer, Larissa; Sampson, Margaret; Bukutu, Cecilia; Schmid, Christopher H.; Nikles, Jane; Tate, Robyn et al. (2016): CONSORT extension for reporting N-of-1 trials (CENT) 2015: explanation and elaboration. In *Journal of clinical epidemiology* 76, pp. 18–46. DOI: 10.1016/j.jclinepi.2015.05.018.

Wernsdorff, Melina von; Loef, Martin; Tuschen-Caffier, Brunna; Schmidt, Stefan (2021): Effects of open-label placebos in clinical trials: a systematic review and meta-analysis. In *Sci Rep* 11 (1). DOI: 10.1038/s41598-021-83148-6.

Yang, Jiabei; Steingrimsson, Jon A.; Schmid, Christopher H. (2021): Sample size calculations for n-of-1 trials. Available online at http://arxiv.org/pdf/2110.08970v1.

Zucker, D. R.; Schmid, C. H.; McIntosh, M. W.; D'Agostino, R. B.; Selker, H. P.; Lau, J. (1997): Combining single patient (N-of-1) trials to estimate population treatment effects

and to evaluate individual patient responses to treatment. In *Journal of clinical epidemiology* 50 (4), pp. 401–410. DOI: 10.1016/S0895-4356(96)00429-5.