Statistical Analysis Plan (SAP)

for

Final Analysis

Full Study Title:	Disentangling pharmacological and expectation effects in antidepressant discontinuation	
Short Study Title:	PHEA	
EudraCT:	n.a.	
NCT-Number:	NCT05191277	
Authors:	A. Ozga. S. Lezius, C. Meißner, C. Warren, Y. Nestoriuc, A. Zapf	
Study Protocol Version:	5.0	
Sponsor's Protocol Code:	PV7151	
Sponsor:	Deutsche Forschungsgemeinschaft (DFG)	
Version / Date:	01-00	17.01.2024
Effective Date:	Date of last signature	
Responsible Bio statistician: *Prof. Dr. Antonia Zapf Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg*	_____ Date	_____ Signature
Responsible Employee (IMBE): *Susanne Lezius Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg*	_____ Date	_____ Signature
Responsible Employee (ISN): *Carina Meißner Helmut Schmidt University / University of the Federal Armed Forces Hamburg Clinical Psychology &*	_____ Date	_____ Signature

University Medical Center Hamburg-Eppendorf, Hamburg, Institute of Systems Neuroscience		
Principal Investigator / Representative of Sponsor: *Prof. Dr. Yvonne Nestoriuc Helmut Schmidt University / University of the Federal Armed Forces Hamburg Clinical Psychology & University Medical Center Hamburg-Eppendorf, Hamburg, Institute of Systems Neuroscience*	_____ Date	_____ Signature

Document History:

Version No.	Date	Chapter / Attachment	Reason for Review / Modification
First draft	09.12.2022	n.a.	n.a.
Second draft	13.02.2023	all	PI comments
First version	23.03.2023	1.3.2, 5.1.2	definition Posner task

Abbreviations

Abbreviation	Definition
AE	Adverse events
AM	Antidepressant medication
AN(C)OVA	Analysis of (co)variance
BAS	Behavioral Approach System
BDI	Beck-Depression-Inventory
BFI	Big Five Inventory
BIS	Behavioral Inhibition System
CONSORT	Consolidated Standards of Reporting Trials
DESS	Discontinuation Related Signs and Symptoms Scale
DFG	Deutsche Forschungsgemeinschaft
DSM	Diagnostic and Statistical Manual of Mental Disorders
EFS	Evaluated for safety set
FAS	Full analysis set
FU1-FU3	Follow-Up 1 to 3
GASE	Generic Assessment of Side Effects
GEEE	Generic rating scale for previous treatment experiences, treatment expectations, and treatment effects
HC	Hidden continuation
HD	Hidden discontinuation
ICC	Intra-class correlation
I/E	Inclusion/Exclusion
IMBE	Institute of Medical Biometry and Epidemiology
ISN	Institute of Systemic Neuroscience
ITT	Intention-to-treat
MADRS	Montgomery-Asperg depression rating scale
MDD	Major depressive disorder
OC	Open continuation
OD	Open continuation
PDI	Pain Disability Index
PHEA	Pharmacological and expectation effects in antidepressant discontinuation
PHQ	Patient Health Questionnaire
PP	Per Protocol
PSS	Perceived Stress Scale
rsfMRI	Resting state functional magnetic resonance imaging
S1, S2	Screening points
SAE	Serious adverse events
SAP	Statistical Analysis Plan
SCID-5-CV	Structured Clinical Interview for DSM-5 – Clinician Version
SSAS	Somatosensory Amplification Scale
STADI	State-Trait-Anxiety-Depression-Scale
SWEMWBS	Short Warwick-Edinburgh Mental Well-Being Scale
t0-t9	Time of measurement
TEX-Q	Treatment Expectation Questionnaire
WHO	World Health Organization

Content

1 Introduction

This Statistical Analysis Plan (SAP) is based on the study protocol version 5.0 of 18.05.2023 and follows the guideline for statistical analysis plans (Gamble, et al., 2017).
Some aspects of the statistical methods and the study design are already described in the study protocol. This SAP aims to further specify the procedures and statistical methods applied during the analysis of the study data.

Analysis will not be conducted but supervised by the Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf.

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

1.1 Background and Rationale

Antidepressant medication (AM) is established as evidence-based, guideline recommended treatment. However, prescription rates are rising (Schwabe et al., 2018) and many patients remain on medication for several years, despite guideline recommendations (Moore et al., 2009). Discontinuation is associated with elevated risk of discontinuation symptoms and recurrence (Henssler et al., 2019) and both are likely to be influenced by dysfunctional expectations (Rutherford et al., 2010). To support patients and physicians in deciding about initiating and discontinuing AM, knowledge about effective and safe ways to discontinue AM is needed. Until today, no study has explored the interplay of patients' expectation (i.e., nocebo effects) and pharmacological effects of AM discontinuation. Disentangling their effects and interaction has important implications for understanding the mechanisms underlying discontinuation symptoms and thus, for the design of future interventions. Hence, we will investigate expectation (high vs. medium expectation induced by verbal instruction) and pharmacological factors (discontinuation of AM vs. continuation) within a balanced open-hidden discontinuation trial.

1.2 Study Objective

This study aims to disentangle pharmacological from expectation effects and their interaction during discontinuation of antidepressants and explore their underlying processes. More specifically the following questions and hypotheses were stated:

- How do treatment and treatment expectation contribute to discontinuation symptom load? Treatment (discontinuation of AM vs. continuation) and treatment expectation (high vs. medium) interact in modulating discontinuation symptom load among remitted major depressive disorder (MDD) patients over the course of the experimental phase.

- In what way can discontinuation symptom load be affected by treatment expectation, i.e. the nocebo-effect, alone? Remitted MDD patients who remain on their antidepressant medication will show a higher discontinuation symptom load with moderate than with high expectation.

- In what way can the discontinuation symptom load be affected by pharmacological factors alone? Remitted MDD patients with moderate treatment expectation will show a higher discontinuation symptom load if antidepressant medication is discontinued versus if antidepressant medication is continued.

- In what way does the discontinuation symptom load differ between high and moderate treatment expectation? Remitted MDD patients who discontinue their antidepressant

medication will show a higher discontinuation symptom load with high than with moderate treatment expectation.

- The relationship between treatment expectation and discontinuation symptom load will vary according to stress ratings, prior side effects of antidepressant medication, prior discontinuation experience, neuroticism, anxiety, somatosensory amplification, and illness framework.

1.3 Study Endpoints

1.3.1 Primary endpoint

Discontinuation symptom load over the course of the experimental phase (i.e., assessments from t2 till t9 will be considered and cover reported symptoms between week 2 until the end of week 13, adjusted for baseline symptoms reported at t1) assessed by the Discontinuation Emergent Signs and Symptoms Scale (DESS; Rosenbaum et al., 1998)

1.3.2 Secondary endpoints

a) Discontinuation symptom load over the clinical observation period (i.e. assessments FU1 till FU3 will be considered and cover reported symptoms between the beginning of week 14 until week 52) assessed by the Discontinuation Emergent Signs and Symptoms Scale (DESS; Rosenbaum et al., 1998)

b) Recurrence (yes or no) over the experimental period, i.e., appearance of a new depressive episode after full, sustained remission of depressive symptoms, will be monitored weekly during run-in and (dis-)continuation phase and biweekly during monitoring phase. Assessments from t1 till t9 will be considered and cover reported symptoms between week 1 until week 13, based on Beck Depression Inventory II and Montgomery Asberg Depression Rating Scale scores (BDI-II, Hautzinger et al., 2009; Montgomery-Asperg Depression Rating Scale [MADRS+], (Davidson et al., 1986; Schmidtke et al., 1988; Williams & Kobak, 2008). If recurrence is suspected (as indicated by BDI-II score >19 or MADRS score >21 over two study visits), corresponding SCID-5-CV sections will be conducted to (dis-)confirm recurrence (Beesdo-Baum et al., 2018; First et al., 2016).

c) First recurrence over the course of the trial, i.e., appearance of a new depressive episode after full, sustained remission of depressive symptoms will be monitored weekly during run-in and (dis-)continuation phase, biweekly during monitoring phase, and at 26, 39, and 52 weeks post-baseline, as determined in a time-to-event analysis. Assessments from t1 till FU3 will be considered and cover reported symptoms between week 1 until week 52 based on Beck Depression Inventory II and Montgomery Asberg Depression Rating Scale scores (BDI-II, Hautzinger et al., 2009; Montgomery-Asperg Depression Rating Scale [MADRS+], (Davidson et al., 1986; Schmidtke et al., 1988; Williams & Kobak, 2008). If recurrence is suspected (as indicated by BDI-II score >19 or MADRS score >21 over two study visits), corresponding SCID-5-CV sections will be conducted to (dis-)confirm recurrence (Beesdo-Baum et al., 2018; First et al., 2016). Calendar week of start of new depressive episode will be examined and recorded.

d) Difference in patient-reported stress between baseline at t1 and 12 weeks later at t9, assessed with the Perceived Stress Scale (PSS-10; Cohen et al 1983; Klein et al 2016)

e) Difference in patient-reported state anxiety between baseline at t1 and 12 weeks later at t9, assessed with the two anxiety-related state subscales of the *State-Trait-Anxiety-Depression-Inventory* (STADI; Laux et al 2016; Renner et al 2018)

f) Attentional and affective processing measured as bias scores (difference between reaction times) of happy versus neutral faces, sad versus neutral faces, and fearful versus neutral faces under high attention to faces in a modified emotional Posner task (Brassen et al., 2011)

1.3.3 Safety endpoints

a) Self- and expert-rated depressive symptoms (BDI-II, Hautzinger et al., 2009; Montgomery-Asperg Depression Rating Scale [MADRS], (Davidson et al., 1986; Schmidtke et al., 1988; Williams & Kobak, 2008) including inspection of suicidality and recurrence
b) Current treatment effects (Generic Rating Scale for Previous Treatment Experiences, Treatment Expectations, and Treatment Effects - Treatment Effects Subscale [GEEEact], Rief et al., 2021)
c) AEs (World Health Organization [WHO], 2000) including life events of subjectively high burden
d) Clinical impression according to psychopathological findings

2 Study Methods

2.1 Trial Design

PHEA is a prospective, randomised, parallel-group, partly blinded, open-hidden discontinuation trial with a 2x2-factorial design. The allocation ratio is 1:1:1:1 for the four groups:

- Treatment discontinuation and high expectation - open discontinuation (OD):
  Patients in this group receive gradually decreasing doses of prescribed AM in encapsulated tablets and will receive full information about their treatment (i.e., high expectation).

- Treatment continuation and high expectation - open continuation (OC):
  Patients in this group remain on prescribed AM in identical encapsulated tablets and will receive full information about their treatment (i.e., high expectation).

- Treatment discontinuation and medium expectation - hidden discontinuation (HD):
  Patients in this group will receive gradually decreasing doses of prescribed AM in encapsulated tablets and will be informed about a 50% chance of discontinuing versus remaining on AM (i.e., moderate expectation).

- Treatment continuation and medium expectation - hidden continuation (HC):
  Patients in this group will remain on prescribed AM in identical encapsulated tablets and will be informed about a 50% chance of discontinuing versus remaining on AM (i.e., moderate expectation).

2.2 Randomization and Blinding

Stratified block randomization was used with variable block sizes.
Stratification variables were

- intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >=24 months. 3:7)
- risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk, 1:1).

The script for the generation of a randomisation list was written by the responsible statistician of the IMBE and transferred to a study-independent employee of the IMBE who created the final randomization list using a new random seed.
The statistic software R (version 4.1.2) (R, 2021) and the package *blockrand* (Snow 2020) was used.

Randomization lists were handed to study staff who have no personal contact to study participants.

The staff responsible for analyses will be blinded with regard to group allocation. The study physician, the study psychologist, as well as supporting student assistants and the study participants are - in the case of assignment to one of the two hidden groups - blinded. Correspondingly, they are unblinded in case of assignment to one of the two open groups. The persons responsible for randomization are unblinded, but have no contact with the study participants. The study physician and the study psychologist are informed that the study participants have been assigned either to the (OD) open-discontinuation group, (OC) open-continuation group, or to one of the two hidden conditions, i.e. either (HD) hidden-discontinuation, or (HC) hidden-continuation. In this case, the study physician and the study psychologist have no knowledge of whether the study participant is in group (HD) or (HC). The information whether the participants belong to the groups (OD) or (OC), or to one of the two hidden groups (HD) or (HC), will be communicated by the study physician or the study psychologist.

2.3 Sample Size

A medium effect size of f=0.22, a significance level of α=0.05, and a power of 80% were assumed. This resulted in 168 patients needed in total (i.e. 42 per group) using an ANOVA model (with main effects and interactions). Furthermore, with an expected drop-out rate of at most 15%, in total 196 patients need to be recruited (i.e. 49 per group).
Sample size calculation was conducted using G*Power (version 3.1.9.2).

2.4 Framework

Test for difference will be applied for all endpoints. The primary endpoint is tested in confirmatory manner. A significance level of 0.05 will be used. All secondary and safety endpoints will be analysed in an exploratory way and thus, no adjustment for multiplicity will be applied.

2.5 Statistical Interim Analyses and Stopping Guidance

No interim analysis is planned.

Criteria for individual study discontinuation will be reviewed during each study visit and, in case of monitoring, during each additional visit. If a participant falls acutely ill (e.g. Covid) during the discontinuation process, a physician may decide to extend the discontinuation process by the duration of the illness (max. 4 weeks). If so, the participant will continue to receive the currently administered dosage of medication from the study team. Measurements will be postponed accordingly and no additional interim measurements will take place. Study participation will be terminated in case of the patient's withdrawal of his/her informed consent, pregnancy, medical or psychological objections by the study physician or psychologist (i.e., severe depressive symptomatology, recurrence, suicidality, other serious or severe AEs, etc.), or insufficient compliance regarding the study participation or the medication. In these cases, the study physician informs about further treatment options and refers to other contact points in an interview. The continuation of the entire study will be questioned by the principal investigator in case of reoccurring AEs that might be causally related to the study treatment regimen, medical or psychological concerns by the study physician or psychologist, no or insufficient study activity (e.g. enrolment rate < 5 per year), or unforeseeable complications that do not justify study continuation. A safety board will monitor the study according to "Guidelines of Good Clinical Practice" (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2015) and will meet regularly. See study protocol for decision path.

2.6 Timing of Outcome Assessments

See Table 1.

Table 1. Schedule of enrolment, intervention and assessments according to SPIRIT-PRO.

	STUDY PERIOD								
	Enrolment		Allocation	Experimental phase			Clinical observation		
Timepoint	**S1**	**S2**	**t0**	**t1-t5**	**t6-t8**	**T9**	**FU1**	**FU2**	**FU 3**
Weeks	**-2**	**-1**	**0**	**1/2/3/4/5**	**7/9/11**	**13**	**26**	**39**	**52**
Enrolment									
Eligibility screen	✓	✓							
Informed consent		✓							
Initial consultation prescribing physician		✓							
Randomization			✓						
Interventions				✓	✓	✓			

Discontinuation versus continuation of AM			✓	✓	✓			
High versus moderate expectation (open versus hidden treatment)			✓	✓	✓			
Assessment								
Primary outcome measures								
Discontinuation symptom load (DESS)		✓	✓	✓	✓			
Secondary outcome measures								
Discontinuation symptom load (DESS)						✓	✓	✓
Recurrence			✓	✓	✓	✓	✓	✓
Psychophysiological stress (PSS-10)*		✓			✓			
State anxiety (STADI State)*		✓			✓			
Attentional, affective processing (Posner task)*					✓			
Possible modulators								
Psychophysiological stress (PSS-10)*		✓						
Side effects of antidepressant medication (GASE*)		✓			✓			
Prior discontinuation experience ($GEEE_{PRE}$*)		✓						
Personality traits (BFI-10)*		✓						
Trait anxiety (STADI State)*		✓						
Somatosensory amplification (SSAS)*		✓						
Illness perception (single item)		✓						
Further assessments								
Optional: rsfMRI		✓	✓ (t1)					
AM blood serum level			✓(t1)		✓			
Trait marker stress*		✓						
Adherence (single item)		✓	✓	✓	✓	✓	✓	✓
Self-reported depressive symptoms (BDI-II)*	✓	✓	✓	✓	✓	✓	✓	✓
Expert-rated depressive symptoms (MADRS)*	✓	✓	✓	✓	✓	✓	✓	✓
Prior discontinuation symptoms ($DESS_{PAST}$)		✓						
Current treatment effects ($GEEE_{ACT}$*)		✓	✓	✓	✓	✓	✓	✓
Expectations (TEX-Q*)		✓	✓	✓	✓			
Expectations ($GEEE_{EXP}$*)		✓	✓	✓	✓	✓	✓	✓
Behavioral inhibition/approach (BIS-BAS scale)*		✓						
Psychopathology (SCID-5-CV interview)*	✓							
Well-being (SWEMWBS)		✓	✓	✓	✓	✓	✓	✓
Depression and Anxiety (PHQ-4)						✓	✓	✓
Subjective Impairment (PDI*)		✓	✓(t1,t5)		✓			
Substance Use & Coping			✓(t1,t5)	✓(t7)	✓			
Screener Warmth & Competence*		✓	✓(t5)		✓			
Adverse events (single safety items)		✓	✓	✓	✓	✓	✓	✓
Suspicions about treatment ($GEEE_{END}$)*					✓			
Demographic* & medical characteristics		✓			✓	✓	✓	✓
Debriefing & close-out								
Debriefing					✓			
Consultation prescribing physician					✓			
Individualized discontinuation plan for continuation groups								✓

Note. S1 = pre-screening; S2 = enrolment; t = assessment timepoint; FU = follow-up; AM = antidepressant medication; DESS = Discontinuation-Emergent Signs and Symptoms Scale; PSS-10 = Perceived Stress Scale, 10 item version; STADI = State-Trait-Anxiety-Depression-Scale, at t9 State only; GASE = Generic Assessment of Side Effects; $GEEE_{PRE}$ = Generic Rating for Treatment Pre-Experiences, Treatment Expectations, and Treatment Effects (previous experiences); BFI-10 = 10-item Big-5 Inventory; SSAS = Somatosensory Amplification Scale; rsfMRI = resting-state functional Magnetic Resonance Imaging; BDI-II = Beck-Depression-Inventory II; MADRS = Montgomery-Asperg Depression Rating Scale; $DESS_{past}$ = Discontinuation-Emergent Signs and Symptoms Scale (previous experiences); $GEEE_{ACT}$ = Generic Rating for Treatment Pre-Experiences, Treatment Expectations, and Treatment Effects (treatment effects); TEX-Q = Treatment Expectation Questionnaire, 15 item version; $GEEE_{EXP}$ = Generic Rating for Treatment Pre-Experiences, Treatment Expectations, and Treatment Effects (treatment expectations); BIS-BAS scale = Behavioral Inhibition and Approach System Scale; SCID-5-CV = Structured Clinical Interview for DSM-5; SWEMWBS = Short Warwick-Edinburgh Mental Well-Being Scale; PHQ-4 = Patient-Health-Questionnaire-4; PDI = Pain Disability Index (adapted to discontinuation symptoms); $GEEE_{END}$ = Generic Rating for Treatment Pre-Experiences, Treatment Expectations, and Treatment Effects (suspicions about treatment); SPIRIT-PRO = Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, SPIRIT-PRO group (2018) Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *Jama*, *319:*483-494. * Part of standardized psychometric test battery.

2.7 Timing of Final Analysis

Analyses might be conducted after completion of the respective study phase, as soon as all data of the respective endpoints are available. Therefore, the analyses of the endpoints concerning the

experimental phase, i.e. the primary analysis concerning discontinuation symptom load over the experimental phase and the analyses concerning the secondary endpoints recurrence over experimental phase, stress, state anxiety, and emotional and attentional processing will be conducted following completion of time point t9 (week 13 after randomization) and the corresponding data cleaning process (i.e. database has been reviewed for completeness and accuracy). According to our first projection, this will happen in the second quarter of 2024.

Analyses concerning the clinical observation phase of the study or the total study time, namely secondary endpoints discontinuation symptom load (during observational phase) and recurrence over the course of the clinical trial will be conducted following completion of FU3 (week 52 after randomization) and data cleaning process (i.e. database has been reviewed for completeness and accuracy). According to our first projection, this will happen in the first quarter of 2025.

3 Statistical Principles

3.1 Confidence Intervals and *P* Values

All applicable statistical tests will be two-sided and will be performed using a 5% significance level. All confidence intervals presented will be 95% and two-sided. Analyses of secondary and safety outcomes will be performed exploratory without adjustment for multiple testing.

3.2 Adherence and Protocol Deviations

Treatment adherence to intake of study medication during the experimental phase is assessed by a single item on the number of days the study medication was taken since the last study visit, as well as via blood analysis. Adherence to intake of study medication is defined as max. 1 omitted pill per week during the discontinuation period.

Major protocol deviations are:

- Non-adherence, i.e., more than 1 omitted pill per week during the discontinuation period, during the experimental phase
- Withdrawal of informed consent
- Pregnancy
- Insufficient compliance regarding requirements for the study participation

Minor protocol deviations are:

- If a participant falls acutely ill (e.g. Covid) during the discontinuation process, a physician may decide to extend the discontinuation process by the duration of the illness (max. 4 weeks). If so, the participant will continue to receive the currently administered dosage of medication from the study team. Measurements will be postponed accordingly; no additional interim measurements will take place.

3.3 Analysis Populations

3.3.1 Intent-to-treat (ITT)

The primary endpoint analysis is based on the intention-to-treat population, i.e. all randomized patients are included, as belonging to their randomization arm, regardless of whether they refused therapy, or whether other protocol violations are known.

3.3.2 Full analysis set (FAS)

For all secondary endpoints the full analysis set is used. FAS is as complete as possible and as close as possible to the ITT population. All available data will be used and all patients with available data will be included as belonging to their randomization arm, regardless of whether they refused therapy, or whether other protocol violations are known.

3.3.3 Per Protocol Population (PP)

The Per Protocol population is a subset of the ITT population and includes only patients who have no major protocol violation.
For the definition of major protocol deviations see Section 3.2.

3.3.4 Evaluated for Safety Set (EFS)

All randomized patients will be included.

4 Trial Population

4.1 Screening Data

Available screening data include:

- Age
- Gender
- Single vs. recurrent depressive episodes

These will be reported.

4.2 Eligibility

The number of ineligible patients randomised, if any, will be reported, with reasons for ineligibility. Ineligible patients are those who do not fulfil inclusion criteria or who show exclusion criteria:

Inclusion criteria

1. Adult patients (18-75 years) with fully remitted MDD, single or recurrent, as main diagnosis confirmed by prescribing physician and SCID-5-CV (Beesdo-Baum et al., 2018; First et al., 2016)
2. Intake of SSRI/SNRI (citalopram: 20-40mg, escitalopram: 10-20mg, sertraline: 75-150mg, venlafaxine: 75-150mg, duloxetine: 60-100mg, paroxetine: 20-40mg) or NaSSA (mirtazapine: 30-45mg)
3. Discontinuation wish by patient, supported by prescribing physician
4. Fulfilment of the German S3 national guideline recommendations for treatment of Major Depressive Disorders to discontinue AM: a) response to AM, b) symptom remission for at least four months (first episode)/ 2 years (2 or more episodes with significant functional impairment) and c) concurrent intake of antidepressant medication (at least 4 weeks on a steady dose)

Exclusion criteria

1. Acute or chronic somatic illness and/or intake of medication which might interfere with depressive disorder, AM or proposed study
2. Acute suicidality, psychotic symptoms, substance abuse or addiction within the last 12 months, current mania, or hypomania confirmed by SCID-5-CV or other psychopathology which might interfere with depressive disorder, AM or proposed study
3. Any history of bipolar disorder or psychosis confirmed by SCID-5-CV
4. Severe stressful life events (e.g., death of a family member) within six months prior to study participation
5. Insufficient German language proficiency.
6. No informed consent.

7. MRI-specific exclusion criteria, if applicable: phobic anxiety, claustrophobia, ferromagnetic implants, etc.

4.3 Recruitment

A CONSORT (Schulz, Altman, & Moher, 2010) flow diagram will be used to summarise the number of patients (per treatment group where appropriate) who were:

- assessed for eligibility at screening
    1. fulfilling I/E-Criteria at screening
    2. not fulfilling I/E-Criteria at screening*
    3. post-hoc violation of I/E-Criteria*
- randomised
- not randomised*
- received the randomised treatment
- did not receive the randomised treatment*
- lost to follow-up*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided

4.4 Withdrawal / Follow-up

Information on the number and level of withdrawals/drop outs, number included in the analysis and the number died will be presented in CONSORT diagram format with numbers and reasons for withdrawal and/or exclusion from analysis given at each visit. Levels of withdrawal are defined as:

- Patient withdraws from treatment, but not from the study, and continues to participate in study visits
- Patient withdraws from study, but all data collected up to date withdrawal can be kept and analyzed
- Patient withdraws from study and withdraws all data

Individual missed assessment time points do not automatically lead to withdrawal from the study.

4.5 Baseline Patient Characteristics

Patients will be described with respect to major baseline demographic parameters and medical characteristic (e.g. age, gender, single vs. recurrent depressive episodes, stress ratings, prior side effects of AM, prior discontinuation experience, neuroticism, anxiety, somatosensory amplification, illness rationale of MDD, AM specifications) both, overall and separately for the four randomised groups.
Categorical data will be summarised by numbers and percentages. Continuous data will be summarised by mean, standard deviation, median, 1st and 3rd quantile, and minimum and maximum. Histograms and boxplots will be used to check the distribution and possible outliers for continuous variables. Number of available observations and number of missing observations will be presented for the treatment groups separately. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted (Senn, 1994).

Baseline characteristics representing possible modulators

- **Stress ratings:**

Will be measured with the *Perceived Stress Scale* (PSS-10; Cohen et al. 1983; Klein et al. 2016), which includes ten items with five response options (0 *'none of the time'* to 5 *'very often'*). Total scores range between 0-40, with higher sum scores indicating higher perceived stress.

- **Prior side effects of AM:**
  Will be measured with the *Generic Assessment of Side Effects Scale* (GASE, Rief et al. 2011). The GASE is a self-report measure including 36 symptom descriptions with 4 severity rating options (0 *'not present'*, 1 *'mild'*, 2 *'moderate'*, 3 *'severe'*). For every reported symptom, participants indicate whether they perceive this side effect as drug-induced or not drug-induced. Medication-attributed total scores (sum of all symptom ratings, causally attributed to AM) range from 0-108, with higher scores indicating higher side effect symptom load.
- **Prior discontinuation experience:**
  Will be assessed with a modified version of the Generic Rating Scale for Treatment Pre-Experiences ($GEEE_{PRE}$, Rief et al. 2021), i.e., two items assessing improvement (0 *'no improvement'* – 10 *'greatest improvement imaginable'*) and worsening (0 *'no worsening* – 10 *'greatest worsening imaginable'*) of patient condition attributed to the most recent discontinuation attempt. A difference of improvement – worsening will be calculated, ranging from -10 to 10, with lower scores indicating a more negative discontinuation experience. Participants with no prior discontinuation experience will be assigned 0.
- **Neuroticism:**
  Will be assessed with the *Emotional Stability subscale of the Brief Big Five Inventory* ([BFI-10], Rammstedt et al. 2007), i.e., two items rated on a five-step scale (1 *'disagree strongly'* to 5 *'agree strongly'*). Higher sum scores indicate higher neuroticism.
- **Anxiety (trait):**
  Will be assessed with the two anxiety-related trait subscales of the *State-Trait-Anxiety-Depression-Inventory* (STADI; Laux et al. 2016, Renner et al. 2018) Each includes 5 statements with four response options (1-4). Total trait-anxiety scores base on sum scores of both anxiety subscales and range between 10 and 40, with higher sum scores indicating higher trait anxiety.
- **Somatosensory amplification:**
  Will be assessed with the *Somatosensory Amplification Scale* ([SSAS], Barsky et al. 1990; Doering et al. 2015*)*, which includes ten items rated on a five-point scale *(1 'not at all true' to 5 'extremely true')*. Total scores range from 10-50 with higher scores indicating more somatosensory amplification.
- **Illness rationale (single item)**:
  i.e., whether patients perceive their depressive disorder as a biologically or a psychologically caused illness, will be assessed with a single item ranging from 0 *'biologically caused'* to 10 *'psychologically caused*'. Lower scores indicate a more biological understanding of depressive disorders.

5 Analysis

5.1 Outcome Definitions

5.1.1 Primary endpoint

Discontinuation symptom load over the course of the experimental phase (week 2 (t2) to week 13 (t9)), will be assessed with the DESS (Rosenbaum et al., 1998), a standardized and widely used self-report measure incorporating 43 discontinuation symptoms of AM. Intensity of each discontinuation symptom will be rated on a 4-point response-scale (0 'not present', 1 'mild', 2 'moderate', 3 'severe'; Tint et al., 2008) weekly (from week 2 to week 5) or biweekly (from week 5 to week 13). Total sum scores range from 0 – 129 with higher scores indicating more pronounced discontinuation symptoms. Discontinuation symptom load will be measured as area under the curve (AUC) over all measurements from week 2 to week 13. In the model, discontinuation symptom load will be adjusted for baseline

discontinuation symptoms during run-in period reported at t1 (i.e., patients remain on their prescribed type and dose of AM but receive their pills in a new-looking, encapsulated form). AUC will be calculated using the trapezoidal rule.

5.1.2 Secondary endpoints

a) Discontinuation symptom load over the clinical observation period (i.e. assessments from FU1 till FU3 will be considered and cover reported symptoms between week 14 until week 52) assessed by the Discontinuation Emergent Signs and Symptoms Scale (DESS; Rosenbaum et al., 1998):
As for the primary endpoint AUC will be calculated from the measurements of FU1, FU2, and FU3.

b) Recurrence (yes / no) during experimental phase, i.e., appearance of a new depressive episode after full, sustained remission of depressive symptoms over the experimental period (i.e., assessments from t1 till t9 will be considered and cover reported symptoms between week 1 until week 13). Recurrence will be monitored weekly during run-in and (dis-)continuation phase and biweekly during monitoring phase, based on Beck Depression Inventory II and Montgomery Asberg Depression Rating Scale scores (BDI-II, Hautzinger et al., 2009; Montgomery-Asperg Depression Rating Scale [MADRS+], (Davidson et al., 1986; Schmidtke et al., 1988; Williams & Kobak, 2008). If recurrence is suspected (as indicated by BDI-II score >19 or MADRS score >21 over two study visits), corresponding SCID-5-CV sections will be conducted to (dis-)confirm recurrence (Beesdo-Baum et al., 2018; First et al., 2016).

   Binary variable first recurrence of depression during experimental phase of trial (i.e. yes/no) will be given at t9.

c) First recurrence over the course of the trial, i.e., appearance of a new depressive episode after full, sustained remission of depressive symptoms will be monitored weekly during run-in and (dis-)continuation phase, biweekly during monitoring phase, and at 26, 39, and 52 weeks post-baseline, as determined in a time-to-event analysis. Assessments from t1 till FU3 will be considered and cover reported symptoms between week 1 until week 52 based on Beck Depression Inventory II and Montgomery Asberg Depression Rating Scale scores (BDI-II, Hautzinger et al., 2009; Montgomery-Asperg Depression Rating Scale [MADRS+], (Davidson et al., 1986; Schmidtke et al., 1988; Williams & Kobak, 2008). If recurrence is suspected (as indicated by BDI-II score >19 or MADRS score >21 over two study visits), corresponding SCID-5-CV sections will be conducted to (dis-)confirm recurrence (Beesdo-Baum et al., 2018; First et al., 2016). Calendar week of start of new depressive episode will be examined and recorded.

d) Change from baseline (t1) in patient-reported stress at t9 (12 weeks after t1), assessed with the Perceived Stress Scale (PSS-10; Cohen et al. 1983; Klein et al. 2016). The PSS-10 includes ten items with five response options (0 *'none of the time'* to 5 *'very often'*). Total scores range between 0-40, with higher sum scores indicating higher perceived stress. Difference scores will be calculated by subtracting total scores at t1 from total scores at t9, resulting in a range from -40 to 40, with higher scores indicating increased stress.

e) Change from baseline (t1) in patient-reported state anxiety at t9 (12 weeks after t1), assessed with the two anxiety-related state subscales of the *State-Trait-Anxiety-Depression-Inventory* (STADI; Laux et al. 2016, Renner et al. 2018). Each includes 5 statements with four response options (1-4). Total state-anxiety scores base on sum scores of both anxiety subscales and range between 10 and 40, with higher sum scores indicating higher trait anxiety. Difference

scores will be calculated by subtracting total scores at t1 from total scores at t9, resulting in a range from -30 to 30, with higher scores indicating increased state-anxiety.

f) Bias scores of the emotional Posner task (Brassen et al., 2011) are the differences between reaction times to emotional faces and neutral.
In short, participants respond as fast as possible to a dot target by button pressing, while neutral, happy, sad, or fearful faces are presented as distractors. Targets are preceded by either spatially-directing cues leading to covert shifts in the attentional focus (i.e., low attentional resources to process distractors) or non-spatial cues, leaving the attentional focus on the faces. We will measure reaction times in milliseconds (ms) for each condition and calculate difference scores between the reaction times to happy and neutral faces, sad and neutral faces, as well as fearful and neutral faces under high attention to faces.

5.2 Missing Data

For the primary endpoint:
Missing values between the measuring time points will be interpolated linearly.
In case of individual termination of study treatment (due to burden of discontinuation symptoms or recurrence), missing data will be imputed as last observation carried forward (LOCF), which should be a conservative approach.
If all values are missing, the result is missing.

For secondary endpoints:

a) The same approach as for the primary endpoints will be applied, i.e. LOCF.

b) No imputation of missing values will be conducted. The analysis is based on the FAS.

c) Patient observations will be censored at the time point where treatment switching is observed, i.e.

- treatment discontinuation if treatment should be continued
- treatment intake if it should be discontinued

Other missing data will not be imputed.

d) No imputation of missing values will be conducted. The analysis is based on the FAS. As sensitivity analysis the analysis will be repeated on multiple imputed data sets.

e) No imputation of missing values will be conducted. The analysis is based on the FAS. As sensitivity analysis the analysis will be repeated on multiple imputed data sets.

f) No imputation of missing values will be conducted. The analysis is based on the FAS.

For multiple imputation a fully conditional specification strategy will be used to generate 20 imputations. The results pooled according to Rubin's rule will be reported.

For the primary endpoint and the secondary endpoint a) the analysis based on imputed data (preserving the ITT population) will be primarily reported.
For all other secondary endpoints the analysis based on the FAS will be primarily reported and analyses based in imputed data are sensitivity analyses.

If other imputation methods are necessary for further additional analyses they will be defined in an additional SAP.

5.3 Efficacy Evaluation

For the primary endpoint the ITT population will be used. For the secondary endpoints the ITT population or FAS will be used as described above.
Mean, standard deviation, median, minimum and maximum for the continuous variables and absolute and relative frequencies for the categorical variables are presented both overall and separately for the randomized groups for each time point. Numbers of missing observations are presented for the randomized groups separately for each time point.

5.3.1 Analysis of Primary Endpoint

A linear model (ANCOVA) will be used with the following variables:

- Dependent variable:
  - AUC value based on DESS scores t2-t9 (see Section 5.1.1)

- Independent variables:
  - Random group as two factors (i.e. 1st factor: treatment discontinuation vs continuation; 2nd factor; high vs. medium expectation) and their interaction

  - The following stratification variables of the randomization process:
    - *intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >24 months)*
    - *risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk)*

  - The corresponding pseudo baseline value (t1) will additionally be included as covariate.

Primary interest lies in the interaction between the two factors describing the random group. The post-hoc comparisons between the subgroups will be addressed exploratively.
If the corresponding p-value (for the interaction) is >0.15 the model will be calculated without the interaction and the independent main effects of the 2 factors will be assessed.

Normal distribution of the residuals will be checked graphically via a histogram of residuals. If the assumption of normally distributed residuals is clearly violated, a sensitivity analysis using the log-transformation for the dependent variable will be performed to judge the robustness of the results.

Adjusted mean values and adjusted effect estimators (mean differences) with corresponding 95% Wald confidence intervals and p-values will be given.

5.3.2 Analysis of Secondary Endpoints

Continuous outcomes

Linear models (ANCOVA) will be used with the following variables:

- Dependent variables:
  - AUC value based on DESS scores week 14 till week 52, i.e., assessed at FU1-FU3
  - Change from baseline (t1) of patient-reported stress scores based on PSS-10 scores at t9 (12 weeks after t1)

  - Change from baseline (t1) of patient-reported state-anxiety scores based on STADI scores at t9 (12 weeks after t1)

- Independent variables:
  - Random group as two factors (i.e. 1st factor: treatment discontinuation vs continuation; 2nd factor; high vs. medium expectation) and their interaction
  - The stratification variables of the randomization process:
    - *intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >24 months)*
    - *risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk)*
  - Concerning secondary endpoint a), the corresponding pseudo baseline value (t1) will additionally be included as covariate.

Primary interest lies in the interaction between the two factors describing the random group. The post-hoc comparisons between the subgroups will be addressed exploratively.
If the corresponding p-value (for the interaction) is >0.15 the model will be calculated without the interaction and the independent main effects of the 2 factors will be assessed.

Normal distribution of the residuals will be checked graphically via a histogram of residuals. If the assumption of normally distributed residuals is clearly violated, a sensitivity analysis using the log-transformation for the dependent variable will be performed to judge the robustness of the results.
Adjusted mean values and adjusted effect estimators (mean differences) with corresponding 95%-Wald confidence intervals and p-values will be given.

The ICC (intra-class correlation) will also be reported, where appropriate.

Linear mixed model (random intercept model) will be used with the following variables:

- Dependent variable:
  - Bias score (emotional versus neutral faces, measured in ms)

- Independent variables (fixed factors):
  - Expectation concerning recurrence of discontinuation symptoms derived from the factors of the random group: high expectation (open discontinuation group) versus moderate expectation (both hidden groups, continuation and discontinuation) versus low expectation (open continuation group)
  - Type of emotion (happy versus sad versus fearful)
  - 3x3 interaction of expectation and emotion
  - The stratification variables of the randomization process:
    - *intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >24 months)*
    - *risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk)*

- Variable used as random effect (intercept)
  - Patient ID

Primary interest lies in the interaction between expectation and emotion. The post-hoc comparisons between the expectation - subgroups regarding the different emotions will be addressed exploratively.
If the corresponding p-value (for the interaction) is >0.15 (i.e. the expectation groups do not differ substantially regarding the different emotions) the model will be calculated without the interaction

and the independent main effects of the 2 factors will be assessed (again with post-hoc comparisons if appropriate).

Adjusted mean values and adjusted effect estimators (mean differences) with corresponding 95%-Wald confidence intervals and p-values will be given.

The ICC (intra-class correlation) will also be reported, where appropriate.

Binary outcome

Binary logistic model will be used with the following variables:

- Dependent variable:
  - Binary variable judging any recurrence of depression during 13 week experimental period (i.e. yes/no) at t9

- Independent variables:
  - Random group as two factors (i.e. 1st factor: treatment discontinuation vs continuation; 2nd factor; high vs. medium expectation) and their interaction
  - The stratification variables of the randomization process:
    - *intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >24 months*),
    - *risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk*)

Primary interest lies in the interaction between the two factors describing the random group. The post-hoc comparisons between the subgroups will be addressed exploratively.
Again, a model without the interaction term of the random group-factors but only the 2 main factors is calculated if the corresponding interaction p-value is above 0.15.

Rates and adjusted effect estimators (odds ratios) with corresponding 95%-Wald confidence intervals and p-values will be given.

Time-to-event outcome

A Cox proportional hazards model will be used for the endpoint first recurrence over the course of the clinical trial. Independent variables are:

- Random group as two factors (i.e. 1st factor: treatment discontinuation vs continuation; 2nd factor; high vs. medium expectation) and their interaction
- The stratification variables:
  - intake duration with 24 months as a marker for long-term intake (<24 months of AM vs. >24 months),
  - risk of developing discontinuation symptoms associated with the antidepressant medication (moderate vs. higher or unknown risk)

Primary interest lies in the interaction between the two factors describing the random group. The post-hoc comparisons between the subgroups will be addressed exploratively.
If the corresponding p-value (for the interaction) is >0.15 the model will be calculated without the interaction and the independent main effects of the 2 factors will be assessed.

The proportional hazards assumption will be checked graphically via Schoenfeld residuals.
Event rates and adjusted effect estimators (hazard ratios) with corresponding 95%-Wald confidence intervals and p-values will be given.

5.4 Sensitivity Analyses

In case of a violation of the normality assumption of the residuals, the primary analysis will be repeated with log transformed dependent variable, where appropriate.
Results of the post-hoc group comparisons will then be presented as adjusted mean factors with corresponding 95%- Wald confidence intervals (resulting from back-transformation of the adjusted mean differences) and p-values will be given.

The primary endpoint and the secondary endpoint a) will additionally be analyzed using the complete case data only, without adjustment for baseline discontinuation symptoms, and in the per protocol population.

Further sensitivity analyses on imputed data might be performed, see Section 5.2 for details. All endpoints might be analysed in the per protocol population.

5.5 Safety Evaluation

All safety endpoints for the treatment groups will be analysed descriptively, i.e. event rates with two-sided 95% confidence intervals will be given.

5.6 Additional Analyses

We assume that the relationship between treatment expectation and discontinuation symptom load will vary according to stress ratings, prior side effects of antidepressant medication, prior discontinuation experience, neuroticism, anxiety, somatosensory amplification, and illness framework. In case of a significant interaction within the primary analysis, individual moderators will be included as 3rd factor in the model and three-way interactions will be analysed.

Saliva and blood samples will be taken as objective markers of AM blood serum level and stress, and optional participation in rsfMRI measurements will be scheduled.

As an objective marker, salivary cortisol awakening response and salivary alpha-amylase activity will be assessed. Salivary analyses will be conducted in the laboratory of the Institute of Medical Psychology and Behavioral Immunobiology, University Hospital Essen, as part of central scientific project Z02 within our CRC 289 Treatment Expectation.

Concerning analysis of the blood samples, we check adherence to study medication intake by comparing AM blood serum levels at t1 and t9 using paired t-tests per respective continuation (no relevant difference expected) or discontinuation group (relevant difference expected). We will assess whether AM blood serum levels lie within therapeutic range at t1 and exploratively examine the relationship between initial AM blood serum levels and discontinuation symptom load reported by the two discontinuation groups with correlation analyses. We hypothesize a positive correlation between initial AM blood serum levels and discontinuation symptom load among discontinuation groups.

In order to identify interindividual differences in effects of expectation on health outcome, brain imaging data on functional and structural connectivity will be collected and contributed to central scientific project Z03 within CRC 289 Treatment Expectation. Functional and structural imaging data from (i) rsfMRI, (ii) 3D-MPRage T1-weighted sequence, and (iii) Diffusion Tensor Imaging (DTI) will be acquired using a 3T MR system (Trio, Siemens Healthcare, Erlangen, Germany at the Institute of Systems Neuroscience, director C. Büchel) with a 64-channel head/neck coil. Standardized MR protocols are provided by Z03, who will later use this data for pooled and meta-analytic analyses.

5.7 Data Problems

5.8 Differences to Trial Protocol

6 Statistical Software

- R 3.4.1 or newer
- SPSS© 22.0 or newer

7 References

Alberts, J., Löwe, B., Glahn, M. A., Petrie, K., Laferton, J., Nestoriuc, Y [Yvonne], & Shedden-Mora, M. (2020). Development of the generic, multidimensional Treatment Expectation Questionnaire (TEX-Q) through systematic literature review, expert surveys and qualitative interviews. BMJ Open, 10(8), e036169. https://doi.org/10.1136/bmjopen-2019-036169

Barsky, A. J [Arthur J.], Wyshak, G., & Klerman, G. L. (1990). The somatosensory amplification scale and its relationship to hypochondriasis. Journal of Psychiatric Research, 24(4), 323–334.

Beesdo-Baum, K [K.], Zaudig, M., & Wittchen, H. U. (2018). Strukturiertes Klinisches Interview für DSM-5®-Störungen Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5® Disorders Clinician Version von Michael B. First, Janet B.W. Williams, Rhonda S. Karg, Robert L. Spitzer. Hogrefe. https://books.google.de/books?id=f34DxQEACAAJ

Brassen, S., Gamer, M., & Büchel, C. (2011). Anterior cingulate activation is related to a positivity bias and emotional stability in successful aging. Biological Psychiatry, 131–137. https://doi.org/10.1016/j.biopsych.2010.10.013

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. Journal of Personality and Social Psychology, 67(2), 319.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. Journal of Health and Social Behavior, 24(4), 385. https://doi.org/10.2307/2136404

Davidson, J., Turnbull, C. D., Strickland, R., Miller, R., & Graves, K. (1986). The Montgomery-Asberg Depression Scale: Reliability and validity. Acta Psychiatr Scand, 73(5), 544–548. https://doi.org/10.1111/j.1600-0447.1986.tb02723.x

Dillmann, U., Nilges, P., Saile, H., & Gerbershagen, H. U. (2011). PDI - Pain Disability Index - deutsche Fassung. https://doi.org/10.23668/PSYCHARCHIVES.4505

Doering, B. K., Nestoriuc, Y [Y.], Barsky, A. J [A. J.], Glaesmer, H [H.], Brähler, E [E.], & Rief, W [W.] (2015). Is somatosensory amplification a risk factor for an increased report of side effects? Reference data from the German general population. Journal of Psychosomatic Research, 79(6), 492–497. https://doi.org/10.1016/j.jpsychores.2015.10.010

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). SCID-5-CV: Structured clinical interview for DSM-5 disorders: Clinician version. American Psychiatric Association Publishing Arlington.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. TRENDS in COGNITIVE SCIENCES, 11(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Dore, C., . . . Loder, E. (2017). Guidelines for the content of statistical analysis plans in clinical trials. JAMA - Journal of the American Medical Association, 318, 2337-2343.

Hautzinger, M., Keller, F., & Kühner, C. (2009). BDI-II: Beck-Depressions-Inventar (2. Aufl.). Pearson.

Haver, A., Akerjordet, K., Caputi, P., Furunes, T., & Magee, C. (2015). Measuring mental well-being: A validation of the Short Warwick-Edinburgh Mental Well-Being Scale in Norwegian and Swedish. Scandinavian Journal of Public Health, 43(7), 721–727. https://doi.org/10.1177/1403494815588862

Henssler, J., Heinz, A., Brandt, L., & Bschor, T. (2019). Antidepressant Withdrawal and Rebound Phenomena. Deutsches Arzteblatt International, 116(20), 355–361. https://doi.org/10.3238/arztebl.2019.0355

Klein, E. M., Brähler, E [Elmar], Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., Wölfling, K., & Beutel, M. E. (2016). The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. BMC Psychiatry, 16, 159. https://doi.org/10.1186/s12888-016-0875-9

Lang, G., & Bachinger, A. (2017). Validation of the German Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a community-based sample of adults in Austria: a bi-factor modelling approach. Journal of Public Health, 25(2), 135–146. https://doi.org/10.1007/s10389-016-0778-8

Laux, L [L.], Hock, M [M.], Bergner-Köther, R [R.], Hodapp, V., & Renner, K.-H [K.-H.]. (2016). STADI: State-Trait-Angst-Depressions-Inventar. In Diagnostische Verfahren in der Psychotherapie (3rd ed., pp. 478–483). Hogrefe. https://fis.uni-bamberg.de/handle/uniba/41475

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H [Heide], Wingenfeld, K., Schneider, A., & Brähler, E [Elmar] (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord, 122(1-2), 86–95. https://doi.org/10.1016/j.jad.2009.06.019

Moore, M., Yuen, H. M., Dunn, N., Mullee, M. A., Maskell, J., & Kendrick, T. (2009). Explaining the rise in antidepressant prescribing: A descriptive study using the general practice research database. BMJ (Clinical Research Ed.), 339, b3999. https://doi.org/10.1136/bmj.b3999

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Renner K-H, Hock M, Bergner-Köther R, Laux L. Differentiating anxiety and depression: the State-Trait Anxiety-Depression Inventory. Cognition & Emotion. 2018;32:1409–23. doi:10.1080/02699931.2016.1266306.

Rief, W [Winfried], Barsky, A. J [Arthur J.], Glombiewski, J. A., Nestoriuc, Y [Yvonne], Glaesmer, H [Heide], & Braehler, E. (2011). Assessing general side effects in clinical trials: Reference data from the

general population. Pharmacoepidemiology and Drug Safety, 20(4), 405–415. https://doi.org/10.1002/pds.2067

Rief, W [Winfried], Nestoriuc, Y [Yvonne], Mueller, E. M., Hermann, C., Schmidt, K., & Bingel, U. (2021). Generic rating scale for previous treatment experiences, treatment expectations, and treatment effects (GEEE). https://doi.org/10.23668/psycharchives.4717

Rosenbaum, J. F., Fava, M., Hoog, S. L., Ascroft, R. C., & Krebs, W. B. (1998). Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. Biological Psychiatry, 44(2), 77–87. https://doi.org/10.1016/S0006-3223(98)00126-7

Rutherford, B. R., Wager, T. D., & Roose, S. P. (2010). Expectancy and the Treatment of Depression: A Review of Experimental Methodology and Effects on Patient Outcome. Current Psychiatry Reviews, 6(1), 1–10. https://doi.org/10.2174/157340010790596571

Schmidtke, A., Fleckenstein, P., Moises, W., & Beckmann, H. (1988). Untersuchungen zur Reliabilität und Validität einer deutschen Version der Montgomery-Asberg Depression-Rating Scale (MADRS) [Studies of the reliability and validity of the German version of the Montgomery-Asberg Depression Rating Scale (MADRS)]. Schweizer Archiv fur Neurologie und Psychiatrie (Zurich, Switzerland : 1985), 139(2), 51–65.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. International Journal of Surgery, 63, 834-840.

Schwabe, U., Paffrath, D., Ludwig, W.-D., & Klauber, J. (2018). Arzneiverordnungs-Report 2018. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-57386-0

Senn, S. (1994). Testing for baseline balance in clinical trials. Statistics in medicine, 13(17), 1715-26.

Snow, G. (2020). blockrand R package. https://CRAN.R-project.org/package=blockrand

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. Health and Quality of Life Outcomes, 7, 15. https://doi.org/10.1186/1477-7525-7-15

Strobel, A., Beauducel, A., Debener, S., & Brocke, B. (2001). Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White. Zeitschrift Für Differentielle Und Diagnostische Psychologie, 22(3), 216–227. https://doi.org/10.1024//0170-1789.22.3.216

Tait, R. C., Pollard, C. A., Margolis, R. B., Duckro, P. N., & Krause, S. J. (1987). The Pain Disability Index: psychometric and validity data. Arch Phys Med Rehabil, 68(7), 438–441.

Williams, J. B. W., & Kobak, K. A. (2008). Development and reliability of a structured interview guide for the Montgomery-Åsberg Depression Rating Scale (SIGMA). British Journal of Psychiatry, 192(1), 52–58. https://doi.org/10.1192/bjp.bp.106.032532

World Health Organization. (2000). Safety monitoring of medicinal products: guidelines for setting up and running a pharmacovigilance centre. http://apps.who.int/medicinedocs/en/d/Jh2934e/15.html