Statistical Analysis Plan (SAP)

for

Final Analysis

Full Study Title:	The Efficacy of Automated Feedback After Internet-based Depression Screening: the German, Three-armed, Randomised Controlled Trial DISCOVER
Short Study Title:	DISCOVER
EudraCT:	NCT04633096
Authors:	Marion Schütt
Study Protocol Version:	Published study protocol (Sikorski et al., 2021)
Sponsor's Protocol Code:	n.a.
Sponsor:	University Medical Center Hamburg-Eppendorf
Version / Date:	V01-0 30-09-2022
Effective Date:	Date of last signature

Responsible Bio statistician:
Prof. Antonia Zapf
University Medical Center Hamburg-Eppendorf, Department of Medical Biometry and Epidemiology

30.09.2022 (Date) — A. Zapf (Signature)

Responsible Employee / Representative of Responsible Employee :
Anne Daubmann
University Medical Center Hamburg-Eppendorf, Department of Medical Biometry and Epidemiology

30.09.2022 (Date) — (Signature)

Principal Investigator / Representative and Sponsor / Representative:
Sebastian Kohlmann, PhD
University Medical Center Hamburg-

30.09.2022 (Date) — S. Kohlmann (Signature)

Eppendorf, Department
of Psychosomatic
Medicine and
Psychotherapy

Document History:

Version No.	Date	Chapter / Attachment	Reason for Review / Modification
0.1	29.06.2022	n.a.	n.a.
0.2	06.07.2022	All chapters	Comments project team
0.3	07.07.2022	All chapters	Comments project team
0.4	20.09.2022	All chapters	Comments project team
0.5	28.09.2022	All chapters	Review cycle in the project team
1.0	30.09.2022		Final

Abbreviations

Abbreviation	Definition
ANCOVA	Analysis of Covariance
Brief IPQ	Brief Illness Perception Questionnaire
CI	Confidence interval
CSSRI	Client Sociodemographic and Service Receipt Inventory
EFS	Evaluated for Safety Set
FAS	Full Analysis Set
GAD-7	Generalized Anxiety Disorder Scale (7 items)
IEC	Independent Ethics Committee
ITT	Intention-To-Treat principle
IRB	Institutional Review Board
LOCF	Last Observation Carried Forward
NMAR	Not Missing At Random
PD	Protocol Deviation
PHQ-9	Patient Health Questionnaire (Depression severity, 9 items)
PP	Per Protocol
SAP	Statistical Analysis Plan
SCID-5-CV	Structured Clinical Interview for DSM-5 Disorders
SSS-8	Somatic Symptom Scale (8 items)
USE	Usefulness Scale
VAS	Visual Analogue Scale

Content

1 Introduction 6
1.1 Background and Rationale 6
1.2 Study Objective 6
1.3 Study Endpoint(s) 6
1.3.1 Primary Endpoint 6
1.3.2 Secondary endpoints 6
1.3.3 Other endpoints 7
1.3.4 Safety endpoints 7
1.3.5 Participants characteristics 7
2 Study Methods 9
2.1 Trial Design 9
2.2 Randomization and Blinding 9
2.3 Sample Size 10
2.4 Framework 10
2.5 Statistical Interim Analyses and Stopping Guidance 10
2.6 Timing of Outcome Assessments 11
2.7 Timing of Final Analysis 11
3 Statistical Principles 12
3.1 Confidence Intervals and *P* Values 12
3.2 Intervention Adherence and Protocol Deviations 12
Intervention Adherence 12
Protocol deviations 12
3.3 Analysis Populations 12
Full Analysis Set (FAS) 12
Per Protocol Population (PP) 12
Evaluated for Safety Set (EFS) 12
4 Trial Population 13
4.1 Screening Data 13
4.2 Eligibility 13
4.3 Recruitment / Withdrawal / Follow-up 13
4.4 Baseline Participant Characteristics 14
5 Analysis 14
5.1 Outcome Definitions 14
Primary Outcome 14
Secondary Outcome 14
Other Outcomes 14

Safety Outcome 14
5.2 Missing Data 14
5.3 Efficacy Evaluation 15
Analysis of Primary Endpoint 15
Analysis of Secondary Endpoints 15
Analysis of other Endpoints 15
Sensitivity Analyses 15
5.4 Safety Evaluation 16
5.5 Additional Analyses 16
Check of Assumptions 16
Drop Out Analysis 16
Control for Blindness of the Participants 16
5.6 Data Challenges 16
5.7 Differences to Trial Protocol 16
5.8 Statistical Software 17
6 References 17

1 Introduction

This Statistical Analysis Plan (SAP) is based on the published study protocol (Sikorski et al., 2021) 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.

The description of the health economic evaluation is not subject of this SAP.

1.1 Background and Rationale

Depression is one of the most disabling disorders worldwide, yet it often remains undetected. One promising approach to address both early detection and disease burden is depression screening followed by direct feedback to participants. Evidence suggests that individuals often seek information regarding mental health on the internet. Thus, internet-based screening with automated feedback has great potential to address individuals with undetected depression.

1.2 Study Objective

The study objective is to determine whether automated feedback after internet-based depression screening reduces depression severity as compared to no feedback.

Primary Hypothesis: The depression severity 6 months after screening is lower in at least one of the two feedback study arms (STANDARD FEEDBACK and/or TAILORED FEEDBACK) as compared to the NO FEEDBACK study arm.
Secondary Hypothesis: The depression severity 6 months after screening is lower in the TAILORED FEEDBACK arm as compared to the STANDARD FEEDBACK arm.

1.3 Study Endpoint(s)

1.3.1 Primary Endpoint

The primary endpoint of the study is the change in self-reported depression severity total score from baseline to six months after randomization, measured via the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Löwe et al., 2004). The PHQ-9 consists of 9 items and is scored on a 4-point Likert scale (0–3), resulting in a total score ranging from 0 to 27 (0 = best, 27 = worst).

1.3.2 Secondary endpoints

- PHQ-9 total score change from baseline to one month after randomization.
- Depression diagnosis by a health care professional, resulting in a binary variable (yes / no), measured by self-report 6 months after randomization.
- Guideline-based depression care, using a self-developed questionnaire (psychotherapy and/or medication), based on the German National Clinical Practice Guideline for Unipolar Depression (DGPPN et al., 2015), resulting in a binary variable (yes / no), measured 6 months after randomization.

- Depression-related health behaviour, using a self-developed questionnaire (information-seeking, seeking social support, self-management, seeking formal help), resulting in a binary variable (yes / no), 6 months after randomization.
- Health-related quality of life will be assessed using the EuroQol-5D (EQ-5D-5L, Ludwig etal., 2018) six months after randomization, change from baseline. The EuroQol-5D index-value is generated using 5 dimensions: mobility, self-care, activity, pain and anxiety (each on a 5-point scale: 1 = best, 5 = worst). In addition, a visual analogue scale (VAS) is assessed.
- Anxiety severity measured via the 7-item Generalized Anxiety Disorder total score change (GAD-7, Spitzer et al., 2006, German version: Löwe et al., 2008) six months after randomization, change from baseline.
- Somatic symptom severity measured via the 8-item Somatic Symptom Scale total score (SSS-8, Gierk et al., 2014) six months after randomization, change from baseline.

1.3.3 Other endpoints

- Illness beliefs regarding depressive symptoms will be measured with a modified version of the well validated Brief Illness Perception questionnaire (Brief IPQ, Broadbent et al., 2006) and analysed per domain / item.
- Major depression diagnosis evaluated with the Depression related modules of the Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV, Beesdo- Baum et al., 2019). Interviews will be conducted via telephone by a trained assessor.Sum score of number of critical life events: Two open questions assessing relevant positive and negative critical life events will be asked by phone 6 months after randomization and summed up.

1.3.4 Safety endpoints

To estimate possible unintended adverse events of the feedback intervention, participants are asked about the occurrence of any negative event that is attributed to the trial with an open question six months after randomization. Adverse events will be categorized by the investigator.

1.3.5 Participants characteristics

Participant characteristics recorded before randomization include
- sociodemographic data (e.g. age, gender, education, family status, rural/urban area living, local residency),
- medical data (diagnosis of and treatments for depression) and
- risk factors for depression onset (e.g. chronic somatic comorbidities, pregnancy, alcohol and nicotine consumption).

Domain	German question	Notes
Rural/urban area living	Wie groß ist die Stadt, in der Sie leben?	
	In welchem Bundesland leben Sie?	
	Wie sind Sie krankenversichert?	
Demographic data	Wie groß sind Sie (ggf. schätzen)?	
	Wie schwer sind Sie (ggf. schätzen)?	
Local residency	Was ist Ihre Muttersprache?	
	Was ist Ihre Staatsangehörigkeit?	
	Würden Sie sich selbst als Migrant/in bzw. Person mit Migrationshintergrund bezeichnen?	
Gender	Was ist Ihr Geschlecht?	
Age	Wie alt sind Sie?	

Family status	Was ist Ihr Familienstand?	
	Wie ist Ihre Wohn-/ Lebenssituation?	
Education	Was ist Ihr höchster Schulabschluss?	Summarized into 3 categories
	Was ist Ihr beruflicher Status?	
	Besteht im Moment oder bestand in den letzten sechs Monaten ein reguläres Arbeitsverhältnis (außer Minijob)?	
Nicotin consumption	Rauchen Sie (Zigaretten, E-Zigaretten, Zigarillos, Pfeife)?	
	Wie viele Zigaretten (bzw. E-Zigaretten, Zigarillos, Pfeifenköpfe, etc.) rauchen Sie am Tag?	
	Seit wie vielen Jahren rauchen Sie (wenn Sie weniger als 1 Jahr rauchen, geben Sie bitte 0 ein)?	
Alcohol consumption, AUDIT-C	Wie oft trinken Sie Alkohol? Wenn Sie an einem Tag Alkohol trinken, wie viele alkoholhaltige Getränke trinken Sie dann typischerweise? Wie oft haben Sie im letzten Jahr an einem Tag 6 oder mehr alkoholische Getränke getrunken?	sum score
Risk factor score	Leiden Sie unter Ängsten? Leiden Sie unter einer Sucht (Drogen, Computerspielsucht, Spielsucht, etc.)? Hatten Sie in der Vergangenheit ein Lebensereignis, das Sie bis heute belastet? Leiden Sie seit mindestens sechs Monaten unter anhaltenden körperlichen Beschwerden (z.B. Rückenschmerzen, Kopfschmerzen, Übelkeit)? Litten Sie im letzten Monat unter Stimmungsschwankungen oder gedrückter Stimmung? Haben Sie (eine) chronische körperliche Erkrankung(en) (z.B. Herzerkrankung, Diabetes, Asthma, etc.)? Fühlen Sie sich von Ihrem sozialen Umfeld unterstüzt? (invert) Wurde bei Ihnen die Diagnose einer anderen psychischen/seelischen Erkrankung gestellt? Haben Sie Familienmitglieder, die unter psychischen Beschwerden leiden (z.B. Depressionen, Ängste, Essstörungen, Suchterkrankungen)? Gibt es in Ihrer Familiengeschichte bekannte Fälle von Suiziden oder Suizidversuchen? Versuchen Sie im Moment schwanger zu werden? Sind Sie momentan schwanger? Haben Sie in den letzten 6 Monaten entbunden? Stillen Sie momentan? Leiden Sie unter prämenstruellen Störungen (PMS)? Befinden Sie sich in Ihrer Menopause?	sum score
Somatic morbidity score	Um welche chronische(n) Erkrankung(en) handelt es sich? Herzerkrankung Diabetes Atemwegserkrankungen (z.B. Asthma, COPD, etc.) Darmerkrankungen Neurologische oder Nervenerkrankungen (z.B. Multiple Sklerose, Schlaganfall, etc.) Schmerzerkrankungen Krebs	sum score

	Rheumatologische Erkrankungen und/oder Gelenkserkrankungen sonstige Erkrankung(en)	
Depression history	Wurde bei Ihnen jemals die Diagnose Depression oder Burnout gestellt?	
	Wer hat die Diagnose erstmalig gestellt und Ihnen vermittelt?	
	Wann waren Sie das letzte Mal aufgrund von Depressionen oder Burnout in Behandlung?	
	Um welche Art der Behandlung handelt(e) es sich?	
	Können Sie sich vorstellen, aktuell an einer Depression zu leiden?	
Comorbidities	Wurde bei Ihnen die Diagnose einer anderen psychischen/seelischen Erkrankung gestellt?	
	Haben Sie (eine) chronische körperliche Erkrankung(en) (z.B. Herzerkrankung, Diabetes, Asthma, etc.)?	
Potential treatment preference	Wenn Sie vermuten würden, an einer Depression zu leiden, welche/n Behandler/in würden Sie als erstes aufsuchen?	

2 Study Methods

2.1 Trial Design

The DISCOVER trial is designed as an internet-based, observer-blinded, stratified randomized controlled clinical trial with three parallel groups (1:1:1), which is conducted nationwide in Germany.

Study arms:

- No feedback: The participants do not get any feedback on their screening result.
- Standard feedback: Participants receive standardized feedback comprising the following four sections 1) the depression screening result, 2) a note to seek diagnostic consultation by a health professional, 3) brief general information on depression and 4) information on depression treatment.
- Tailored feedback: Participants receive standard feedback tailored to their individual symptom profile, illness perceptions and preferences. Details regarding the tailored feedback arm are described in the published study protocol.

2.2 Randomization and Blinding

Randomisation is based on a computer-generated randomization sequence (1:1:1 allocation ratio), which was conducted by an independent researcher of the Department of Medical Biometry and Epidemiology and is not accessible to any other study team member. The sequence consists of permuted blocks of randomly arranged sizes (6, 9 and 12) and is stratified by baseline depression severity (moderate: PHQ-9 ≥ 10 and ≤ 14 points; severe: PHQ-9 ≥ 15 points). Allocation is performed by a computerized system, ensuring allocation concealment.
Individuals who participate multiple times are automatically allocated to the same study arm as before. This process is ensured by a privacy-preserving record linkage service which identifies double entries based on personal data.

Participants know their allocation due to the nature of the intervention but are kept unaware of trial hypotheses to minimise expectancy bias. The research staff assessing outcomes in the telephone interviews are blind to the allocation at any time.

2.3 Sample Size

Based on the results of the preceding DEPSCREEN-INFO trial (Löwe et al., 2016), the study is powered to detect a small mean difference (Cohen's $f = 0.118$) in the primary outcome (depression severity) in any pairwise comparison between all three study arms. The calculation is based on a global one-way ANCOVA adjusted for baseline depression severity, with an alpha of 0.05 (two-sided) and a power of 80%. It results in a needed sample size of $n = 233$ participants per group (PASS, 2008). To allow for an estimated drop out of 35% (c.f. Christensen et al., 2009), 358 participants per group are recruited (1074 in total).

2.4 Framework

The DISCOVER study is a superiority trial, testing pairwise comparisons between three study arms. All hypotheses are formulated two-sided and test for differences. To show superiority, we additionally look at the point estimate and the two-sided 95% confidence interval.

2.5 Statistical Interim Analyses and Stopping Guidance

No interim analysis is conducted.

2.6 Timing of Outcome Assessments

Table 1. Measures and assessment time points for endpoints

Measures	T0	T1	T2	T3
Primary outcome				
Depression severity, PHQ-9	x		x	x[a]
Secondary outcomes / other outcomes				
Depression diagnosis by a health care professional				x
Guideline-based depression care (psychotherapy and/or medication)				x
Depression-related health behaviour (e.g. information-seeking)				x
Anxiety severity, GAD-7	x			x
Somatic symptom severity, SSS-8	x			x
Health-related quality of life, EQ-5D-5L	x			x
Illness beliefs, Brief IPQ	x		x	x
Intervention adherence	x			
Critical life events				x[b]
Depression diagnosis, SCID		x[b]		x[b]
Adverse events				x[b]
Characteristics				
Sociodemographic data	x			
Medical data	x			
Risk factors for depression onset	x			

Note. T0 = before randomisation (The randomization takes place immediately after the T0 assessments are performed); T1 = 2 days after randomisation; T2 = 1-month follow-up, T3 = 6-months follow-up; PHQ-9 = Patient Health Questionnaire-9; EQ-5D-5L = EuroQol-5D 5-L; GAD-7 = Generalized Anxiety Disorder-7; SSS-8 = Somatic Symptom Scale-8; SCID = Structured Clinical Interview for DSM-5 Disorders; Brief IPQ = Brief Illness Perception Questionnaire.

[a] Primary outcome; [b] Measures assessed via telephone interview.

2.7 Timing of Final Analysis

The final analysis of the DISCOVER trial takes place as soon as the final visit of the last participant is completed, the data are collected, the queries are processed, and the database is locked. According to our current milestone plan, the final data transfer will take place early October 2022. This is followed by the final analysis.

3 Statistical Principles

3.1 Confidence Intervals and *P* Values

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

3.2 Intervention Adherence and Protocol Deviations

Intervention Adherence

Adherence will be defined using feedback reading time ≥ 15 sec (yes / no) and/or download of feedback form.

Protocol deviations

Major protocol deviations are defined as follows:

- Participants that did not adhere to or did not receive the intervention (feedback reading time < 15 sec or no download of feedback form)
- Participants that have participated repeatedly for whatever reason
- Participants that report having participated in the study before
- Participants that report not having answered the survey seriously
- Participants with a baseline survey completion time < 2 min
- Participants with an invalid email address

3.3 Analysis Populations

Full Analysis Set (FAS)

The primary analysis is based on the full analysis set (FAS). It is as complete as possible and as close as possible to the Intention-To-Treat (ITT) principle which includes all randomized participants, as belonging to their randomization arm, regardless of whether they received the feedback or not, or whether other protocol violations are known. For the FAS at least a valid baseline and one valid post-baseline value of the primary outcome needs to be available.

Per Protocol Population (PP)

The Per Protocol population is a subset of the FAS and includes only participants who have no major protocol violation (see chapter 3.2).

Evaluated for Safety Set (EFS)

All randomized participants who were provided directly after the randomization with feedback or a 'thank you' note (no feedback) will be included into the Evaluated for Safety (EFS) set. Only participants who were reached by phone 6 months after randomization could be asked regarding possible adverse events.

4 Trial Population

4.1 Screening Data

Reporting of the number of screened participants (participants who started T0) and the number of screening failures.
Patients with an invalid email address will be regarded as not eligible after randomization. Due to technical reasons the validity could only be checked after randomization.

If available, baseline participant characteristics of the screening failures and reasons for exclusion are analysed using summary statistics.
Categorical data are summarized by absolute and relative frequencies. Continuous data are summarized by mean, standard deviation, median, first and third quartile, minimum, and maximum. These measures are presented for the total screening population.

4.2 Eligibility

The following eligibility criteria according to the study protocol are assessed within a self-report online survey at T0 directly before randomization:

Participants are required to
- be aged 18 years or above,
- show an indication for at least moderate depression (PHQ-9 ≥ 10 points),
- provide contact details,
- be willing to give informed consent.

Furthermore, participants who clicked on the website, gave informed consent and entered their personal data were seen as having
- sufficient German language proficiency,
- have internet access and
- have sufficient computer/internet literacy.

Participants are excluded
- if they were diagnosed with depression within the past 12 months or
- if they currently are or were receiving depression treatment within the past 12 months.

Further, in some cases technical circumstances occurred before randomization and impeded randomization.
Due to technical reasons the validity of the email address could only be checked after randomization. Participants with an invalid email address will be regarded as not eligible after randomization.

4.3 Recruitment / Withdrawal / Follow-up

A CONSORT flow diagram is used to summarize the number of participants who were:

- assessed for eligibility at screening (started T0)
- eligible at screening and completed T0
- ineligible at screening*
- eligible and randomized
- eligible but not randomized*
- received the randomized feedback arm or no feedback

- did not receive the randomized feedback / no feedback
- lost to follow-up*
- discontinued the intervention*
- randomized and included in the primary analysis
- randomized and excluded from the primary analysis*

*reasons are provided

4.4 Baseline Participant Characteristics

Summary statistics are given for the baseline participant characteristics (see chapter 1.3).
Categorical data are summarized by absolute and relative frequencies. Continuous data are summarized by mean, standard deviation, median, inter-quartile range, minimum, and maximum. These measures are presented for the total screening population and by study group.

5 Analysis

5.1 Outcome Definitions

Primary Outcome

- PHQ-9 total score (9 items) change (T3 (6 months after randomization) – T0 (baseline))

Secondary Outcome

The following secondary endpoints are continuous outcomes. For these, difference from baseline (T0) is calculated for 1 month (T2; PHQ-9 only) and 6 months after randomization (T3, all outcomes):

- PHQ-9 total score (9 items)
- GAD-7 total score (7 items)
- SSS-8 total score (8 items)
- EQ-5D-5L (index-value, VAS)

Other Outcomes

The following other endpoints are continuous outcomes:

- Brief IPQ single items (difference from baseline (T0) is calculated for 1 month (T2) and 6 months after randomization (T3))
- Critical life events at T3 (sum score for positive and negative life events, respectively)

The following other endpoints are binary outcomes. For these, no difference to baseline (T0) is calculated:

- SCID at T1 and T3

Safety Outcome

- Adverse events at T3

5.2 Missing Data

The multilevel modelling approach limits the bias when handling missing data even in the case of not missing at random (NMAR).
As a sensitivity analysis missing values are imputed by different approaches for the primary analysis. In the FAS population no baseline value is missing as per definition. In case of missing follow-up values,

we conduct last observation carried forward (LOCF) and multiple imputation. In the LOCF approach, missing values are replaced by the previous valid value. The imputation model in the multiple imputation approach includes all baseline characteristics of participants and all variables of the primary outcome analysis. From the baseline values of the secondary outcomes, we include as many variables in the imputation model as possible. For the imputation of missing follow-up values, 100 imputed data sets are generated, and the results are combined using Rubin's Rules.

5.3 Efficacy Evaluation

Analysis of Primary Endpoint

A multilevel model incorporating the participants as a random term is applied to the repeated measures in the same participant, including the factor group and the baseline value as a covariate. The primary analysis is performed within the framework of this model as an ANCOVA of the PHQ-9 change scores (T0 to T3-difference), with subsequent pairwise comparisons of interventions by test of the corresponding contrasts. The primary analysis is performed in the respective FAS population. The pairwise comparisons are done if the global test of group (H_0: all means are equal) could be rejected. Each test will be performed at a two-sided level of alpha = 0.05. This closed testing principle will ensure a family-wise error level of 5%.

Absolute, and relative frequencies of missing observations are presented both overall and separately for the randomized groups.

Analysis of Secondary Endpoints

The secondary endpoint analyses are performed in the respective FAS population. All secondary endpoints are compared and statistically assessed for descriptive purposes and not in a confirmatory sense. The aim of the analysis is an exploratory data analysis, not hypothesis testing or generation of evidence for the efficacy. No attempt is made to adjust for the p-values for multiple testing.
Mean, standard deviation, first and third quartile, minimum and maximum for the continuous variables and absolute and relative frequencies for the categorical and binary variables are presented both overall and separately for the randomized groups for each time point. Number of missing observations are presented for the randomized groups separately for each time point. All summary tables are structured with a column for each group (tailored feedback, standard feedback, no feedback) and a total column.
Multilevel linear models with differences from baseline as outcome, incorporating the participants as a random term, are applied to the repeated measures in the same participant, including the factor group, time point, and the baseline value as a covariate. The interaction between time and the group is determined. If it is not significant, it is eliminated from the model. Adjusted means with 95% confidence intervals (CI) and p-values are reported.

Analysis of other Endpoints

Other endpoints will be analysed, depending on the level of scale, as defined for the secondary endpoints. See chapter 5.3. The SCID will be analysed descriptively using summary statistics.

Sensitivity Analyses

The primary endpoint analysis is repeated within the PP population.

Further for the primary endpoint a multilevel linear model with change from baseline as outcome incorporating the participants as a random term are applied to the repeated measures in the same

participant, including the factor group, time point, and the baseline value as a covariate. Adjusted means with 95% CI and p-values are reported.
See section 5.2 for sensitivity analyses regarding missing value imputation.

In addition, a subgroup analysis regarding the baseline depression severity (moderate: PHQ-9 ≥ 10 and ≤ 14 points; severe: PHQ-9 ≥ 15 points) is planned. The model defined for the primary analysis of the PHQ-9 change score will be extended by the factor depression severity (moderate / severe). The interaction of the feedback-group and the depression severity will be investigated. In case the p-value of the interaction is < 0.05 we consider that the depression severity at baseline is influencing the PHQ-9 effect.

5.4 Safety Evaluation

All adverse events will be analysed in the EFS using summary statistics and, if applicable, analysed using logistic regression.

5.5 Additional Analyses

Check of Assumptions

Normality of residuals and linearity
The assumptions regarding normality of residuals, and linearity between the independent variables and the dependent variable are examined graphically. Partial residual plots are used to examine linearity. Residual plots and quantile-quantile plots are used to evaluate normality of residuals. In case of unmet assumptions, the Box-Cox transformation is applied in order to find out which power transformation is reasonable in terms of not violating the assumption.

Drop Out Analysis

If applicable, we perform a drop out analysis. Therefore, we use a binary variable (drop out: yes vs. no) as a dependent variable in a logistic regression. The baseline characteristics as well as the baseline values of primary and secondary outcomes are used as effects.

Control for Blindness of the Participants

Steps to control for blindness include the following: after every interview, assessors are (a) instructed to document if participants have disclosed their randomisation status and (b) asked to guess the study arm. After study closure, this guess is compared with the actual status and Cohen's kappa is computed to identify whether hit rates differ from what can be expected from chance.

5.6 Data Challenges

Due to the internet format different scenarios can threaten data validity. Therefore, before the analysis a pre-processing of the data was performed to eliminate invalid entries. Pre-processing was performed by the statisticians, if applicable, and by the principal investigator, if checking of personal data was required. Repeated entries of the same participant were excluded. In case of randomized participants only the first entry was used. In case of not randomized participants the entry with the most information was used.

5.7 Differences to Trial Protocol

In contrast to originally stated in the study protocol, the USE will not be analysed. A change of the USE questionnaires occurred during the study.

Furthermore, website use will not be analyzed as a variable on its own, but will be used for the evaluation of major protocol deviations and intervention adherence (see 3.2).

5.8 Statistical Software

- SAS© 9.4 or newer

6 References

- Sikorski F, König H-H, Wegscheider K, Zapf A, Löwe B, Kohlmann S. The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. Internet Interventions 25. 2021; 100435.
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline E9. Note for guidance on statistical principles for clinical trials (CPMP/ICH/363/96).
- CONSORT Statement (Consolidated Standards of Reporting Trials)
- Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-43