Statistical Analysis Plan

GET.FEEDBACK.GP

Enhancing the Clinical Effectiveness of Depression Screening Using Patient-targeted Feedback in General Practices

Registered at clinicaltrials.gov NCT03988985 with the title 'Enhancing the Clinical Effectiveness of Depression Screening Using Patient-targeted Feedback in General Practices: The GET.FEEDBACK.GP Multicentre Randomized Controlled Trial' (clinicaltrials.gov)

Version: 2.0

Date: November 21, 2022

Author: Mika Tode, A. Fierenz, Antonia Zapf

Reviewers: Eik Vettorazzi

Institution: UKE

Document:	Statistical Analysis Plan for Final Analysis
Sponsor Project Code:	GET.FEEDBACK.GP
Study No.	Registered at clinicaltrials.gov NCT03988985 with the title 'Enhancing the Clinical Effectiveness of Depression Screening Using Patient-targeted Feedback in General Practices: The GET.FEEDBACK.GP Multicentre Randomized Controlled Trial' (clinicaltrials.gov)
Study short title:	Enhancing the Clinical Effectiveness of Depression Screening Using Patient-targeted Feedback in General Practices
Version / date:	**2.0** / **November 21, 2022** based on study protocol version draft 2, dated February 31, 2019
Author:	Mika Tode, A. Fierenz, Antonia Zapf
Sponsor:	University Medical Center Hamburg-Eppendorf
Statistician:	Prof. Dr. Antonia Zapf Department of Medical Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg

Table of Contents

Version history

Version	Date	Section number changed	Description/Changes made
Draft 1	16.01.19		First draft
Draft 2	31.01.19		Minor changes to first draft
Draft 3	03.06.19		Minor changes to second draft
Draft 4	12.06.19		Minor changes to third draft
1.0	26.06.19		Finalisation including minor changes
1.1	01.07.19		Minor changes, ready for signature
1.2	11.11.22	5.2.3	Prespecified subgroup added
2.0	29.11.22		Final and included signatures

Signature Page

Prof. Dr. Bernd Löwe

Principal Investigator, sponsor delegated person

Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg

21.11.2022

Date

Signature

Prof. Dr. Antonia Zapf

Trial Statistician

Department of Medical Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg

29.11.2022

Date

Signature

Mika Tode, A. Fierenz, Antonia Zapf

SAP authors

Department of Medical Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg

29.11.2022

Date

Signature

Abbreviations

CEAC	cost-efficiency-acceptancy curve
EFS	Evalued for Safety Set
GP	General Practitioner
SAP	Statistical Analysis Plan
ICER	incremental cost effectiveness ratio
IEC	Independent Ethics Committee
IRB	Institutional Review Board
ITT	Intention-to-treat
PP	Per Protocol
SAP	Statistical Analysis Plan

1 Introduction

This Statistical Analysis Plan (SAP) is based on the study protocol version draft 2 of February 31, 2019 and follows the guideline for statistical analysis plans (Gamble, et al., 2017)
Some points of the statistical methods and of the study design are already described in the study protocol. This Statistical Analysis Plan aims to further specify the procedures and statistical methods applied during the final analysis of the study data.

1.1 Background and rationale

Depressive disorders belong to the most frequent and grave medical conditions. In Germany's general population approximately one in ten citizens is affected by them, means an absolute number of 5 to 6 million adults (1-year prevalence: 9.3%). Those affected lose 8.3 quality-adjusted life years on average and the health costs in Germany add up to a total 4.6 billion € per year. To cure depressive disorder as the most frequent mental disorder an uncomplicated and at the same time effective strategy is needed. This applies mainly for the primary medical care, as about every sixth patient in a general practitioner's (GP) office suffers from a depressive disorder. Roughly half of all depressive disorders are diagnosed and treated by a GP. However, depressive disorders are still not detected and adequately diagnosed in about 50% of all cases. The GET.FEEDBACK.GP intervention addresses this very problem. Showed by the preliminary study called DEPSCREEN-INFO the implementation of patient-oriented feedback additional to the depression screening can evidentially improve mental health. A systematic analysis on the efficacy of depression screenings in primary medical care has never been conducted before in Germany.

1.2 Objectives

This trial is conducted to primarily determine the efficacy of a widely available feedback intervention after depression screening for depressive symptomology in six months as well as costs, depression severity after 1 and 12 months, diagnosis of depression (MINI), anxiety (GAD-7), somatic symptom severity (SSS-8) and quality-adjusted life years as secondary endpoints. Primary hypothesis: patient-oriented electronic depression feedback (including GP feedback) leads to a lower depression severity (PHQ-9) after 6 months compared to both GP feedback and no feedback at all. Secondary hypothesis: patient-oriented feedback after depression screening is more cost efficient after 6 months compared to both GP feedback and no feedback at all caused by a reduction of health costs (CSSRI). Furthermore, the influence

of Germany's first implementation of a depression screening on routine care of primary medical patients is evaluated.

The primary and secondary endpoints are in detail (Table 1 in the approval):

	Measurement			
Instruments	**Baseline**	**1-month follow-up**	**6-months follow-up**	**12-months follow-up**
Primary Endpoint				
Depression severity (PHQ-9)	x	x	x	x
Secondary Endpoint				
Clinical scales (Anxiety severity, GAD-7, somatic symptom severity, SSS-8, quality of life, EQ5D)	x	x	x	x
Patient knowledge and behavior regarding guideline recommendations (e.g., adherence, coping, search for information)			x	x
Health care use (e.g., doctor consultations, medication) (CSSRI)			x	x
Mini-International Neuropsychiatric Interview (MINI)		x	x	x
Patient activation (PAM13-D)			x	
Satisfaction and acceptance of screening and feedback (USE)			x	
Open questions (e.g., strategies in response to a diagnosis)				x
Medical record (i. e., diagnosis, treatment)	x		x	
Demographic information	x			

2 Study Methods

2.1 Trial design

The trial will be conducted as a multicentric, three-arm, observer-blinded, randomised control study under GP conditions. After depression screening using the PHQ-9 questionnaire eligible patients will be randomly put into one of three equally large study arms by specially programmed tablets administrated by a study nurse. Both patient-oriented plus GP feedback and feedback for the GP only after depression screening belong to the intervention whereas the no feedback arm belongs to the control group. At baseline patients directly fill in a questionnaire. The endpoints for the 1-, 6- and 12-months follow-up are collected via telephone interviews, which was already tested successfully in the preliminary study DEPSCREEN-INFO. The primary endpoint of this study is a change in the Patient health questionnaire-9 (PHQ-9) score for depression severity between baseline and the 6-months follow-up. The PHQ-9 is one of the most frequently used and best validated questionnaire for self-rating depressive conditions. It has the distinction of combining high sensitivity and specificity parameters and it is sensitive for changes considering depression treatment.

2.2 Randomization

A randomising sequence is created (boxes for 97 patients with probabilities for each condition derived from estimated prevalences and two test entries – 999 total boxes) not being

accessible for neither study nurse nor GP at any given point. After a patient has finished the baseline measurement, reached a PHQ-9 score higher than 9, randomisation happens electronically (allocation concealment). Patients are stratified by their depression severity (moderate and severe). The study is not cluster-randomised, as within every single GP's office each of the three conditions is realised and therefore an individual randomisation of patients is feasible.

2.3 Sample size

The sample size calculation is based on the results of the preliminary study called DEPSCREEN-INFO, which showed an effect of Cohen's d = 0.27. In this study a lower effect of d = 0.25 for the intervention group compared to the other groups is supposed. The sample size was calculated for an ANOVA based on the F-test. Because of the closed test principle, in case of a significant result in this overall test, the pairwise comparisons can be performed without adjustment of the type-I error. To detect the underlying effect size through an F-test with a significance level of $\alpha = 0.05$, a power of $1-\beta = 0.8$ and group effects 0 and 0.25 a sample size of n = 233 is needed in every group (balanced design). This adds up to a total sample size for the actual enrolment of N = 699 (computations were made with PASS 2008). The estimated prevalence of depressive disorders in a GP's office is 15%. 50% of all participants of the depression screening are not going to be eligible to further participate in the study as they achieve a lower PHQ-9 score than 10. Based on previous work a loss of 35% of included patients at the time of the 6-months follow-up is expected. Therefore, 1074 patients have to be included in the study in order to roughly maintain 699 patients for the primary analysis. Five study centres with each 10 GP's offices therefore have to include at least 14320 patients in the study. Thus, in every GP's office 287 patients (rounded from 286.4) will be screened on average. 30 patients will be asked to participate at each GP's office per day whereas only 15 of them consent and are eligible for the study. The total recruitment length then adds up to 14350 / (15*5) = 192 recruitment days if patients will be recruited simultaneously in all 5 centres. Considering weekends leads to 192 * (7/5) = 269 days or 39 weeks of recruitment. To take unforeseen recruitment problems into consideration the total length is extended to 12 months.

2.4 Framework

GET.FEEDBACK.GP is planned to show superiority of the "patient-oriented and GP feedback" arm compared to "no feedback" or "GP feedback only" regarding the primary endpoint change in PHQ-9 score.

2.5 Statistical interim analyses and stopping guidance

No interim analysis is planned. The trial will be stopped early if the Data Safety and Monitoring Board (DSMB) existing of the scientific advisory board sees any relation between participation or a certain trial condition and negative events (Serious adverse events, SAE). The scientific advisory board, principal investigator and co-project manager will discuss this option before deciding.

2.6 Timing of final analysis

The final analysis of the primary endpoint (depression severity after 6 months) will take place after the database has been reviewed for completeness and accuracy and is determined by the Steering Committee to be complete for database lock. According to our projection, this will happen after about 3 years. The results will be reported according to the CONSORT 2010 Guideline (Schulz, Altman, & Moher, 2010)

2.7 Timing of outcome assessments

The outcome and endpoints of each patient are found out in a 1-, 6- and 12-months follow-up after the patient's baseline recruitment via telephone interviews (see Section 1.2).

3 Statistical Principles

3.1 Confidence intervals and *P* values

The level of statistical significance is 5% two-sided. Because of the closed testing principle, the pairwise comparisons of the three groups do not have to be adjusted for multiplicity if the overall test yields a significant result. Accordingly, all applicable statistical tests will be two-sided and will be performed using a 5% significance level. Analyses of secondary outcomes will be performed without adjustment for multiplicity. All confidence intervals presented will be 95% and two-sided.

3.2 Adherence

As the intervention consists only of the feedback, the concept of adherence is not applicable.

3.3 Analysis populations

3.3.1 Intention to treat Population (ITT)

The primary analysis population is the ITT (intention to treat) population. The ITT population consists of all patients randomized.

3.3.2 Per Protocol population (PP)

The Per Protocol population includes all patients randomized who have no major protocol violation.

Major protocol violations are any unapproved changes in the research study design and/or procedures that are within the investigator's control and not in accordance with the IEC/IRB - approved protocol that may affect the participant's rights, safety or well-being, or the completeness, accuracy and reliability of the study data.

Major protocol violation includes ineligible participants who were included in the trial by mistake, and those for whom the intervention or other procedure differed from that outlined in the protocol, or failure of consent process.

3.3.3 Evaluated for Safety Set (EFS)

All randomized patients will be included into the Evaluated for Safety (EFS) set.

4 Trial Population

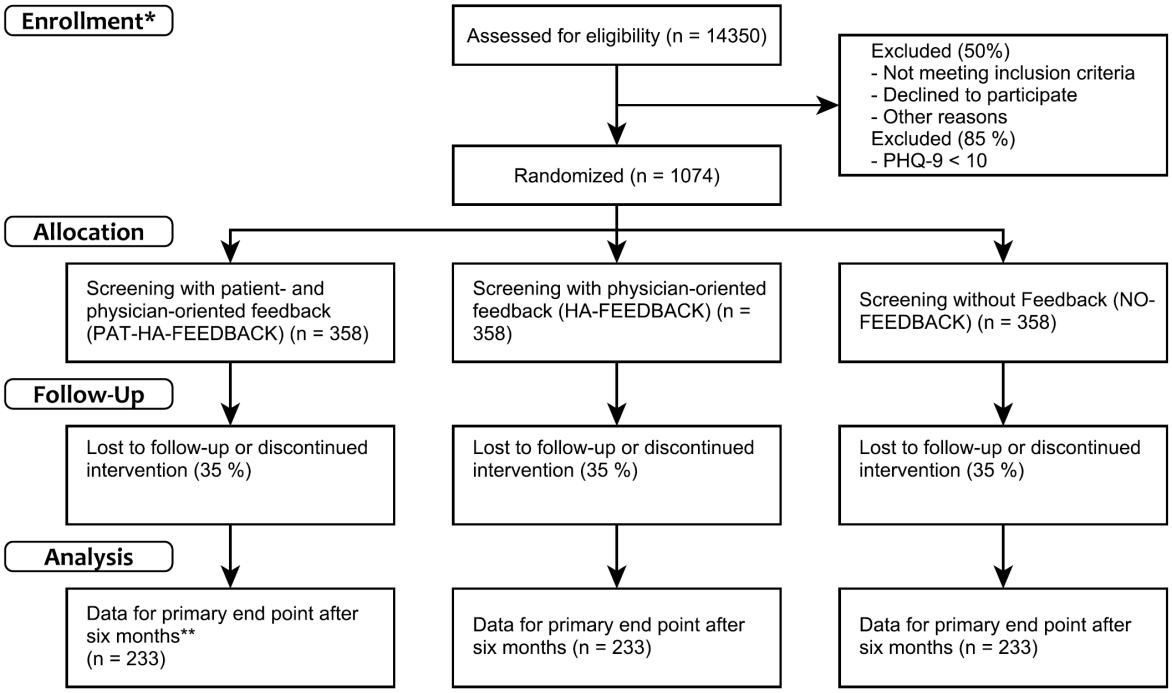
4.1 Screening data

The screening data will be collected using the PHQ-9 questionnaire which evaluates a patient's depression severity.

4.2 Eligibility

A sample is eligible if the individual visits a GP's office with personal contact to the GP, is 18 years old or older, is acceptably proficient in German, consents to participate in the study and is suspected to have a depressive disorder, i.e. a PHQ-9 score greater than 9.
A sample is excluded if the individual has a life threatening health status, suffers from a severe somatic or/and psychological disorder that needs urgent treatment and is not currently diagnosed or treated for depressive disorder or shows an acute suicidal tendency as assessed with PHQ-9 Item 9. In addition, the sample will be excluded if the individual suffers from severe cognitive or/and visual difficulties, is not being able to fill out questionnaires or leaves no contact details.

4.3 Recruitment



* Flow diagram according to Consort Statement for transparent reporting of randomized clinical trials.
** End points after 1 and 12 months are secondary and were not part of the determination of sample size.

4.4 Withdrawal/follow-up

An already accepted patient is withdrawn from the study if the patient withdraws his consent. The study nurse records these events and reports them to the principal investigator. There are three follow-ups planned. The 6-months follow-up is the main one collecting data for both primary and secondary endpoints. The 1-month and 12 months follow-up are only collecting data for secondary endpoints. It is estimated that 35% of the included patients in every group will be lost for the 6-months follow-up. In case of acute suicidality during the follow-up telephone interviews the interviewers will be advised to stop the interview and contact professional help by use of a second phone.

4.5 Baseline patient characteristics

Every patient is invited to the study via the study nurse. The patient consents to participate in the study and receives a depression screening.

The baseline patient characteristics will be summarised in an electronic case-report form (eCRF). As descriptive measures absolute and relative frequencies (for categorical variables), median and quartiles (for skewed-distributed variables), and mean and standard deviations (for variables with assumed normal distribution) are calculated. Appropriate graphics will be used for illustration of the results.

5 Analysis

5.1 Outcome definitions

Primary outcome: depression severity 6 months after intervention measured by the PHQ-9 score.

5.2 Analysis methods

5.2.1 Analysis of primary endpoint

The primary outcome of all randomised patients will be analysed regarding the intention-to-treat principle and according to the memorandum of the DFG about good clinical practice all study results are going to be documented and published.

Primary outcome: depression severity (PHQ-9 score) 6 months after intervention as change from baseline. The primary outcome is going to be analysed via a mixed effects model including random effects patients and GP practices nested in centres and fixed factors for group and depression severity. The baseline comparisons will be represented by statistical contrasts. The use of the closure test ensures a familywise significance level of 5% if the p-value of the overall test is below 0.05.

5.2.2 Analysis of secondary endpoint

The secondary outcomes are health care use (e.g., visiting a doctor, medication), direct and indirect costs, depression severity after one and 12 months (PHQ-9), diagnosis of depressive disorder according to MINI structured interview, anxiety (GAD-7), somatic symptom severity (SSS-8) and quality-adjusted life years (QALYs). Secondary endpoints will be analysed with the data collected in the 1-month, 6-month, and 12-months follow-up like the primary endpoint. Adjusted and non-adjusted means, effect sizes (Cohen's d) and confidence intervals will be reported.

5.2.3 Additional analysis

As additional exploratory analyses, subgroup analyses based on several demographic and clinical values are conducted w.r.t. change of PHQ-9 score. The aim of theses analyses is to identify subpopulations, which benefit more or less from the intervention. These analyses shall give a first starting point for future studies. Missing values will not be imputed, but the implicit MAR assumption of the analysis model will be used.

The presented p-values are of descriptive nature only (two-sided tests at unadjusted significace level of 5%). The analyses are conducted in the FAS.

For the analyses on a potential interaction between the intervention and age (continuous, and categorical: young [≤first tercile] vs. old [>third tercile]), sex (men vs. woman), depression history (previous episode vs. no history of depression), other mental health comorbidity (mental health comorbidity vs. mental health comorbidity), city size (ordinal), education (≤ 10 years in school vs. > 10 years in school), living situation (living together vs. alone), insurance status (private vs. statutory health insurance), reason for GP visit (psychological vs. physical), migration background (migration background vs. no migration background), quality of life at baseline (continuous and dichotomized with median as cut-off point; based on the visual analogue scale of the EQ-5D), anxiety disorder (yes vs. no), substance disorder (yes vs. no), chronic physical disease (yes vs. no), persistent somatic symptoms (yes vs. no), mood swings (yes vs. no), social support (yes vs no), pregnancy in female patients (yes vs. no) and breast feeding in female patients (yes vs. no) on the change in PHQ-9 score from baseline to month 6, the regression models and their parameters are defined analogously to the primary analyses but additional effects for demographic and clinical values and its interaction with the treatment are added to each model.

5.3 Missing data

The mixed effects model as chosen model class is comparatively robust regarding missing data. In addition an analysis using multiple imputation of missing data in baseline variables will be performed as sensitivity analysis.

5.4 Harms

There is no evidence that feedback concerning depression severity causes suicidality or self-harm; it rather seems that questions of suicidality lower the risk of suicide. To guarantee a safe handling with suicidality any evidence of acute suicidality leads to the immediate information of the GP in every of the three study conditions (including the NO-FEEDBACK condition). If acute suicidality is confirmed by the GP a locally coordinated emergency plan in cooperation with the GPs and centres of psychiatry or rather psychosomatic medicine is introduced. In case of suicidal patients blinding is abolished, treatment is introduced and these patients are excluded from the study. Main risks and obstacles during performing of the study are negative events for the patients. For that reason a suitable monitoring is installed. Monitoring and supervision is being controlled by a specialised company (e.g. Clinical Trials Center (CTC) North). A data handling plan agreeing the DFG guidelines and the Federal Data

Protection Act is going to be published beforehand and ensures data availability, transparency and repeatability. Furthermore, the study follows international guidelines such as ICH-GCP (CHMP, 2016) as well as the guidelines and SOPs of the University Medical Centre Hamburg-Eppendorf.

5.5 Statistical software

- STATA 14 or newer
- R 3.4.1 or newer
- SPSS 22.0 or newer

6 References

CHMP. (2016). GUIDELINE FOR GOOD CLINICAL PRACTICE E6(R2) by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *European Medicines Agency, 6*, 1-70.

Committee for Medicinal Products for Human Use (CHMP). (2011). Guideline on missing data in confirmatory clinical trials. *London: European Medicines {|ldots}, 44*, 1-12. doi:10.2307/2290157

European Medicines Agency. (9 1998). *EMA/CPMP/ICH/363/96 - ICH Topic E9 Statistical Principles for Clinical Trials.* Von http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf abgerufen

European Medicines Agency. (1998). Statistical Principles for Clinical Trials. Von http://www.ich.org/fileadmin/PublicWebSite/ICHProducts/Guidelines/Efficacy/E9/Step4/E9Guideline.pdf abgerufen

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, 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. doi:10.1001/jama.2017.18556

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. doi:10.1016/j.ijsu.2010.09.006