Southampton Clinical Trials Unit

Patient-reported outcome measures for monitoring primary care patients with depression (PROMDEP)

Statistical Analysis Plan

Trial name:	Patient-reported outcome measures for monitoring primary care patients with depression (PROMDEP)
Trial registration number:	ISRCTN no: 17299295. Registered 1st October 2018.

UKCRC Registered Clinical Trials Units

Insert logos as per protocol, eg CRUK, NCRN, LRF, BHF etc.

Protocol title and version number:	https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-020-04344-9
SAP version number:	1
SAP date:	6 July 2021

To be approved and reviewed by:

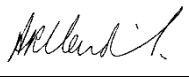
	Name	**Signature**	**Date**
Chief Investigator	Professor Tony Kendrick		06/07/2021
Trial Statistician (author)	Dr Beth Stuart		06/07/2021

Table of Contents

1 Introduction

1.1 Purpose of SAP

This statistical analysis plan (SAP) describes in detail the methods that will be used to analyse the data collected as part of the PROMDEP trial. This will form the basis of the final trial publication. The final analysis will follow the SAP to ensure that the analyses are conducted in a scientifically valid manner and to avoid post hoc decisions which may affect the interpretation of the statistical analysis. Any deviations from the SAP will be detailed in the final report.

1.2 Trial background and rationale (from ISRCTN registration)

The researchers want to look at whether giving personal feedback to people being treated for depression might help them get better more quickly. One way of doing this is by using patient reported outcome measures (or 'PROMs') which involve patients filling out questionnaires to record their symptoms of depression and feeding back the questionnaire results to the health professionals looking after them, at follow-up appointments. Some benefit for patients from reduced depression has been shown to result from monitoring their progress with PROMs, at least in specialist psychological therapy and mental health settings. In a previous study in general practices in southern England between 2014 and 2016, lower levels of depression symptoms were found at 12 weeks follow-up among patients who used PROMs at follow-up assessment, suggesting that completing them may improve the outcome of depression treatment for patients. However, this approach has not yet been researched properly in UK general practices. General practice is the setting in which most people with depression are treated in the UK, so it's important to test whether PROMs can be helpful in that setting.

1.3 Objectives (from published protocol[1])

The objectives of the study are as follows:

To carry out a cluster-randomised, controlled, parallel group trial that will compare (i) getting patients to complete the PHQ-9, which is used as a patient-reported outcome measure (PROM) in their consultations with General Medical Practitioners (GPs) or Nurse Practitioners (NPs) treating them for depression, with (ii) usual Practitioner care, uninformed by PHQ-9 scores

To motivate and train participating practitioners to reflect on the best use of the PHQ-9, thereby improving the practitioner's capability to interpret symptom scores, taking into account patients' responses to open-ended global enquiries, their level of functioning, past history, and social context, including life events and difficulties

To provide patients in the intervention arm with written feedback on their PHQ-9 scores, including a 'traffic light' indication of the level of severity of their depression, a 100-manikin representation of the proportion of people in the population with that level of depression, and a brief list of evidence-based treatments relevant to the level of severity, which they will be asked to discuss with their GP/NP

To follow up participants for 26 weeks, with research assessments at 12 and 26 weeks

To determine the primary outcome of depressive symptoms on the Beck Depression Inventory, 2nd edition (BDI-II), at 12 weeks follow-up

To examine secondary outcomes including depressive symptoms on the BDI-II at 26 weeks, and social functioning, quality of life, and changes in drug treatment and referrals, at both 12 and 26 weeks

1.4 Definition of endpoints

1.4.1 Definition of primary endpoint

Symptom score on the Beck Depression Inventory second edition (BDI-II) for the current level of depression at the 12-week follow-up.

1.4.2 Definition of secondary endpoints

- BDI-II score at 26 weeks
- Scores at both 12 and 26 weeks on the Work & Social Adjustment Scale for social functioning
- Modified version of the Medical Informant Satisfaction Scale MISS at 26 weeks to measure patient satisfaction over the follow-up period.

1.5 Analysis principles

All analyses will be reported according to CONSORT 2010 extension to cluster randomised trials[2] and Southampton Clinical Trials Unit (SCTU) standard operating procedure (SOP) on planning, implementing and reporting statistical analyses (CTU/SOP/5058).

2 Design considerations

2.1 Description of trial design

This is a pragmatic multicentre cluster randomised superiority trial.

2.2 Trial power and sample size

In the PROMDEP feasibility trial[3], we found the mean BDI-II score at baseline was 24.0, and the standard deviation (SD) was 10.0. At the 12-week follow-up, based on the results of the feasibility study, we anticipate a mean of 14.0 in the intervention group and 17.0 in the control group. This gives a mean difference of 3.0 on the BDI-II, which is an effect size of 0.3 SDs and agrees with the findings of Knaup et al's systematic review for the expected effects of combined practitioner and patient feedback of PROMs[4]. The difference of 3.0 points is 17.6% of the control group's score of 17.0 at 12 weeks, and therefore, this score is just above the MCID of 17% for the BDI-II estimated from research with patients[5]. The anticipated potential benefit would therefore be small but clinically significant.

We aim to recruit a mean of six patients per practice. We assume an intra-cluster correlation coefficient (ICC) of 0.03 (from the feasibility study). At the level of 5% significance, to have 90% power to detect a difference between 14.0 and 17.0 on the BDI-II we need 235 patients analysed per group. Given a cluster size of six, the cluster design effect will be 1.15, meaning we need 270 per group. We assume a 20% loss to follow-up at 12 weeks, so the total sample size needed will be 270 × 2/0.8, which is a total of 676 patients recruited from 113 practices across the three recruitment centres (around Southampton, UCL and Liverpool).

2.3 Following discussion with the study steering committees and the funder, the sample size was amended to 554 participants. This allows for a recruitment imbalance of 1.8:1 in favour of the intervention arm but also for a correlation between baseline and follow-up scores of 0.50. with a deflation factor of 1-p^2. Randomisation details

Randomisation of practices is by computerised sequence generation, and minimisation with a random element using three factors to avoid imbalance between the two arms: practice size (large vs small), location (urban/suburban vs rural), and centre (Southampton vs Liverpool vs UCL).

2.4 Timing of planned analyses

2.4.1 Interim analyses and early stopping

No interim analysis is planned and no pre-specified stopping rules have been established.

2.4.2 Final analysis

End of study is defined as when the last patient has had their last data collected, cleaned and verified.

3 Statistical considerations

3.1 Definition of analysis populations

3.1.1 Intention-to-treat analysis population

This population includes all randomised practices and all patients recruited within them regardless of treatment compliance. All summaries and analysis will be on the modified ITT population unless otherwise specified.

3.1.2 Per-protocol analysis population

There is no pre-planned per protocol analysis.

3.2 Analysis software

SAS v9.4 or higher, or Stata v15.1 or higher will be used for all analyses.

3.3 Methods for handling data

3.3.1 Withdrawal from trial

All data up until the point of patient withdrawal from the trial will be used in analyses unless the patient withdrew consent and does not wish for the data already collected prior to withdrawal to be used for the trial.

If a practice withdraws from the trial, no further patients will be recruited. All data on patients collected until that point will be used and any patients recruited will continue to be followed up in accordance with the trial schedule.

3.3.2 Missing data

The primary analysis will be of complete cases.

If more than 2 items in the BDI II or the GAD7 have missing values, the total score will be missing. If one or two items are missing, the score will be imputed with the mean of the non-missing scores before summing.

We will examine the structure and pattern of missing data and, if appropriate, will present a sensitivity analysis based on data imputed using a chained equations multiple imputation model. The imputation model would include the outcome measure, baseline value of the outcome, randomisation group, clustering by practice and all covariates included in the analysis model (see below)

3.3.3 Outliers

No methods will be used to handle outliers in the data, except in the regression models. If outliers are found then firstly the source data will be checked. If the source data is correct, then a sensitivity analysis will be performed excluding them from the analysis.

3.3.4 Assumption checking and alternative methods

Assumptions for linear regression models (linearity, normality, homoscedasticity) will be checked using scatter plots of standardized residuals against fitted values, and qq plots. If linear models are not appropriate a log-linear transformation will be used.

3.4 Definition of key derived variables

The Beck Depression Inventory, second edition BDI-II is a 21-item self-report instrument that uses DSM-IV criteria[6]. It has been established as a valid and reliable instrument for depression screening in the general population and is widely used in depression trials. It takes approximately 5 minutes to complete. Each item is scored from 0 to 3, and a total score of 0–13 is considered minimal range, 14–19 is mild, 20–28 is moderate, and 29–63 is severe.

The Generalised Anxiety Disorder Assessment 7-item version[7] is a self-report measure of anxiety symptoms. Each item is scored from 0-3 with a total possible score ranging from 0-21. Higher scores indicate more severe symptoms.

The Work and Social Adjustment Scale (WASA) assesses problems in functioning with work, home management, social leisure activities, private leisure activities and family and relationships, all on 0 to 8 scales[8]. It has been shown to be a sensitive, reliable and valid measure of impaired functioning and is used routinely in IAPT psychological therapy settings as well as in research studies in a variety of settings.

The 29-item 'Medical Interview Satisfaction Scale' (MISS-29) was developed in the USA to assess patient satisfaction with individual doctor-patient consultations and has been shown to be valid and reliable in UK primary care[9]. We will adapt it to rate patient satisfaction at the 26-week follow-up, asking patients to look back over their consultations with GPs/NPs over the entire 26-week period.

3.5 General principles for reporting and analysis

The following general principles for reporting and analysis will be used:

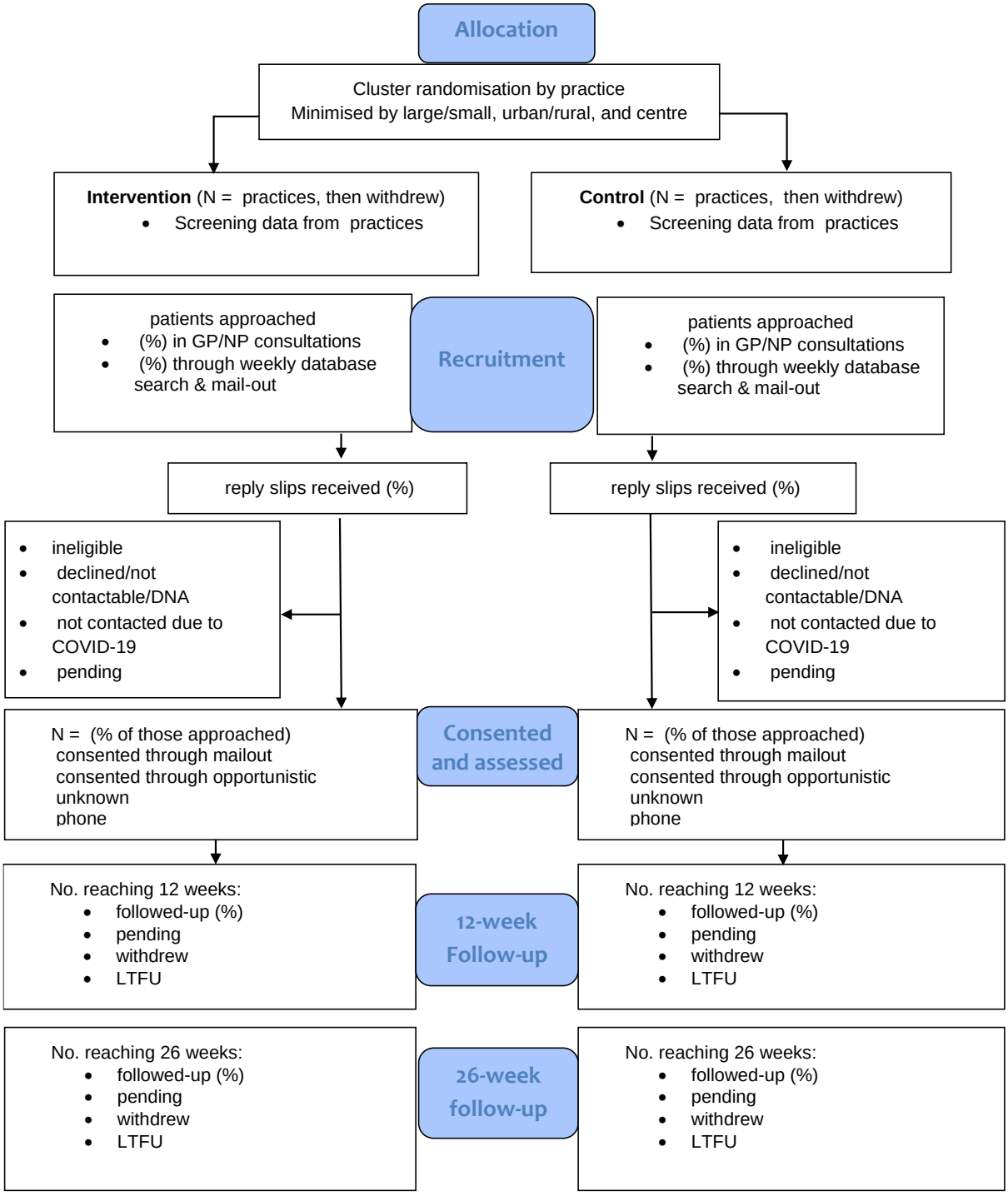
- 5% two-sided level of statistical significance, with corresponding 95% confidence intervals presented where applicable.
- No adjustments for multiplicity are planned.
- Summary statistics will include either mean, standard deviation, and/or median, interquartile range.
- Treatment groups will be labelled in the tables as Intervention Group and Control Group accordingly, and a total column will be included in tables were applicable.

4 Planned analyses and reporting

4.1 Disposition of the study population

A CONSORT diagram will be produced showing a clear account of all practices and patients who entered the trial- see below for an example figure. Withdrawal information including the primary reasons of discontinuation will be summarised and presented by where this is known.

PROMDEP RCT Consort diagram



4.2 Protocol deviations

A listing of all Major or Potential/Serious Breach (Major protocol deviations with potential to affect patient safety/data) and Potential/Serious Breach (with actual affect to patient safety/data, Major/Potential/Serious Breach of GCP guidelines or consistent non-compliance by site) will be produced (by patient and site where applicable)

4.3 Baseline and demographic characteristics

Summary statistics will be produced and presented by group for demographic and baseline characteristics but no comparisons will be undertaken, rather the clinical importance of any imbalance will be noted. If there are imbalances of clinical importance in variables not listed in 4.4 below, we will control for these as covariates in the analyses.

4.4 Primary endpoint

The primary outcome, that is, the differences at 12 weeks between intervention and controls in depression as measured by the BDI-II, will be analysed using a linear mixed model, adjusting for duration of depression, past history of depression, age, gender, marital status, no. of dependents, ethnic group, education level, economic position (employment status), and accommodation status, baseline depression score, anxiety score, and clustering, including practice as a random effect. These covariates have been chosen based on their known relationship with the outcome from previous literature.

The model will use all the observed data and makes the assumption that missing BDI-II scores are missing completely at random.

4.5 Secondary endpoints

Analysis of secondary outcomes, BDI-II at 26 weeks, social functioning and patient satisfaction, will also be conducted using linear regression for continuous outcomes and logistic regression for dichotomous outcomes, again adjusting for socio-demographics, baseline depression, anxiety, and clustering, including practice as a random effect.

4.6 Additional analyses

No subgroup analyses are planned. Any post-hoc analyses will be exploratory only. Health economic analyses will be undertaken and a separate Health Economics Analysis Plan will be prepared.

In accordance with the CONSORT recommendations for cluster randomised trials, we will also report the ICC for the primary outcome.

Due to the COVID-19 pandemic and subsequent lockdown period, it is possible that there may be changes to the key outcomes unrelated to randomisation group. We will therefore look at the scores in each arm in the pre-, peri- and post-COVID periods in the whole study population. We will use descriptive statistics and graphical representations to explore any trends and aim to control for any time varying effect on outcomes in a sensitivity analysis

4.7 Safety reporting

A listing of serious adverse events (SAEs) will be provided for all related/unrelated SAEs. If required, a summary table will also be presented.

5 References

1 Kendrick T, Moore M, Leydon G, *et al.* Patient-reported outcome measures for monitoring primary care patients with depression (PROMDEP): Study protocol for a randomised controlled trial. *Trials* 2020; **21**:441.

2 Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. *BMJ* 2012; **345**. doi:10.1136/bmj.e5661.

3 Kendrick T, Stuart B, Leydon GM, *et al.* Patient-reported outcome measures for monitoring primary care patients with depression: PROMDEP feasibility randomised trial. *BMJ Open* 2017; **7**:e015266.

4 Knaup C, Koesters M, Schoefer D, *et al.* Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. Br. J. Psychiatry. 2009; **195**:15–22.

5 Button KS, Kounali D, Thomas L, *et al.* Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychol Med* 2015; **45**:3269–79.

6 Beck AT, Steer RA BG. No Title. *Man Beck Depress Invent* 1996; **San Antoni**.

7 Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med* 2006; **166**:1092–7.

8 Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *Br J Psychiatry* 2002; **180**:461–4.

9 Meakin R, Weinman J. The 'Medical Interview Satisfaction Scale' (MISS-21) adapted for British general practice. *Fam Pract* 2002; **19**:257–63.