

Randomised control trial of a digital depression management application: juli

Protocol and Analysis plan

Background

Each year, millions of people are diagnosed with depression and seek additional support via digital applications. It can be overwhelming learning how to manage and unclear what can make symptoms better or worse. The digital health application juli, aims to support people with depression via a number of evidence based approaches including mood tracking, medication reminders, positive affect journaling, data visualisation of sleep, activity, exercise, heart rate variability and heuristic behavioural activation recommendations about how to improve these parameters (Katon et al., 1996; Firth et al., 2017).

Trial design and methods

Patients with depression who are interested in using a digital application for support and behavioural activation will be individually randomised between use of the juli app and treatment as usual. The primary outcome will be the PHQ-8 questionnaire that measures depressive symptoms at 8 weeks (Kroenke et al., 2009).

Trial Intervention

Use of juli app vs treatment as usual (plus attention placebo control app).

juli app: Participants will be prompted to open the app each day via an automated alert. They will be asked to rate how they are feeling on a scale using 5 emojis and via a circumplex model with mood on the x-axis and energy on the y-axis. The app will gather information via Apple HealthKit; sleep, activity, workouts, menstrual cycle, heart rate variability. It will present this data to the participant and show associations with mood. It will make recommendations about these parameters to guide healthy behaviours via behavioural activation. The app has a medication reminder function that can be set by the participants to improve medication adherence. Participants are also encouraged to engage in positive affect journaling via the app.

Attention placebo control app: Participants will be prompted to open the app each day via an automated alert and rate how they are feeling on a scale using 5 emojis.

Objectives

The primary objective is to investigate whether the juli app is effective in reducing the severity of depressive symptoms as measured by the PHQ-8 (compared to treatment as usual). The secondary objective is to investigate health related quality of life at 8 weeks using the SF-12.

Trial duration per participant

8 weeks.

Participant locations

Global.

Main inclusion criteria

1) depressive disorder with PHQ-8>4 at baseline, 2) age 18 to 65 (inclusive), 3) English speakers, 4) have an iPhone.

Participant flow

Potential participants will be recruited via online advertisements and support groups. Interested participants will follow a link through to the online patient information sheet. If they wish to participate they will follow a link from the patient information sheet to the online consent form. Following consent they will be directed to download the application. After installing the app they will complete a baseline PHQ-8 and SF-12. If they score <5 on PHQ-8 they will not be entered into the trial, but can continue to use the app. If they score >4 on the PHQ-8 they will be randomised to use the juli app or the attention placebo control app.

Randomisation

Block randomisation will be conducted within the application using automated code, with random block sizes between 4 and 8.

Data collection

All outcome measures will be collected from within the juli app, there is no face-to-face or remote contact between the research team and the participant.

Concept	Measure	Baseline	Week 2	Week 4	Week 6	Week 8
sociodemographics	Bespoke questionnaire	X				
Depression symptoms	PHQ-8	X	X	X	X	X
Health related quality of life	SF-12	X		X		X
User experience	MAUQ					X
Sleep	Time asleep each night	Daily				
Sleep	Time in bed each night	Daily				
Activity	Daily steps	Daily				
Heart rate variability	Heart rate variability	Daily				
Menstrual cycle	Period and predicted fertile window	Monthly				
Workouts	Daily workout time	Daily				
Wellbeing	Likert scale	Daily				
Mood/energy	Circumplex model	Daily				
Medication adherence	Yes/no	Daily				
Journal	Free text	Daily				

Weather	Temperature/ cloud cover/rain	Daily
Air pollution	Air quality index	Daily
Pollen count	Weeds/trees/ grasses	Daily

Abbreviations

PHQ-8 Patient Health Questionnaire 8 item

SF-12 Short Form 12 item

MAUQ mHealth App Usability Questionnaire

Baseline measures

Participants will answer the following questions: age, gender, was depression diagnosed by a physician? How long have you had depression? Do you see a physician regularly about your depression? Do you take regular medication?

They will complete the PHQ-8 and the SF-12.

Follow-up

Daily: wellbeing, mood/energy, medication adherence, journal (juli app only).

Week 2: PHQ-8.

Week 4: PHQ-8, SF-12.

Week 6: PHQ-8.

Week 8: PHQ-8, SF-12, MAUQ.

Statistical analyses

We will follow the Consolidated Standards of Reporting Trials (CONSORT) guidelines in reporting and analysing our data (Schulz et al., 2010).

Flow of participants through the trial

CONSORT diagrams will be produced illustrating the case-finding process and participant throughput in the follow up phase of the study. We will report the following information:

Number of participants recruited via different sources (Google advert, Facebook advert, Twitter, Instagram, Reddit, patient support groups)

Number accessing patient information form

Number completing consent form

Number completing baseline PHQ-8

Number who met eligibility criteria

Number who dropped out after completing baseline information

Number who dropped out after randomisation

Number completing outcome measures at 2, 4 and 8 weeks

Number actively using the app and completing outcome measures at 2, 4 and 8 weeks

Number withdrawing from the trial at 2, 4 and 8 weeks

Number continuing to use the app after 8 weeks

Comparability of groups at baseline

Participants in the two treatment groups will be described separately on the following variables:

Age
Gender
Depression duration
Physician contact
Antidepressant medication
Depression severity (PHQ-8)
Health Status (mental and physical component scores of SF-12)

Numbers (%) will be given for binary and categorical variables, and means and standard deviations (SD) or, medians and inter-quartiles ranges (IQR) for continuous variables, as appropriate. There will be no tests of statistical significance or confidence intervals calculated between randomised groups on any baseline variable.

Description of losses to follow up

The number (%) of participants lost to follow-up at 2, 4 and 8 weeks post-randomisation will be reported for both treatment groups and compared using a chi-square test. Factors associated with 'missingness' of key trial outcomes (e.g., PHQ-8 score) within the dataset at all follow-up points will be explored and described.

Analysis of primary outcome

Within our overall objective, we wish to test the hypotheses that severity of depressive episodes can be reduced using juli. The primary analysis will concentrate on severity of depression.

We have chosen the PHQ-8 at 8 weeks as the primary outcome. The PHQ-9 is a widely used self-completed depression scale which is recommended by the Common Measures in Mental Health Science Governance board (including the National Institute for Mental Health, USA and Wellcome, UK) as one of a core list of research questionnaires that should be used by funded researchers. The PHQ-9 closely matches the DSM-IV criteria for major depressive episode and has been found to have a better sensitivity to change than other measures of depression, such as the Hamilton Rating Scale for Depression and the Beck Depression Inventory (Löwe et al., 2004). The PHQ-8 is preferred in studies where patient contact is remote (such as via digital technologies or via telephone) as it drops the question about suicidality, whilst retaining similar psychometric properties. Research indicates that the deletion of this question has little effect because this question is least frequently endorsed item on the PHQ-9, as such the PHQ-8 has identical scoring thresholds for depression severity (Kroenke et al., 2009; Shin et al., 2019).

The primary outcome is the PHQ-8 score at 8 weeks. The primary analysis will therefore be the difference in total PHQ-8 score at 8 weeks between control and intervention groups. This will be estimated with a linear regression model in which the baseline PHQ-8 will be an adjustment.

We will also carry out a sensitivity analysis using linear regression with adjustment for any variables that are not balanced at baseline (age, gender, depression duration, physician contact, antidepressant medication, SF-12, PHQ-8). We will investigate how robust the findings are to the model specification by testing whether the results are consistent with those produced by a Poisson model.

Analysis of secondary outcomes

The following secondary analyses will be conducted adjusting for the baseline measure of the outcome variable. We will also adjust for any variables that show imbalance at baseline as a sensitivity analysis:

- PHQ-8 score as continuous outcomes at 2, 4, 6 and 8 weeks in a repeated measures analysis
- PHQ-8 score as a binary outcome where remission is defined as scoring <10 on the PHQ-8 at 2, 4, 6 and 8 weeks in a repeated measures analysis
- Difference in SF-12 physical and mental component scores at 8 weeks
- SF-12 physical and mental component scores at 4 and 8 weeks in a repeated measures analysis

Choice of regression model will depend on the outcome. Distributional assumptions appropriate for positive continuous outcomes will be investigated including the log-normal or Poisson distributions. The SF-12 is usually modelled using linear regression as this is often a good fit to the data. Logistic regression will be used for the analysis of binary outcome data.

Approach to dealing with missing outcome data

We will carry out sensitivity analyses to investigate the possible impact of missing data (White et al., 2011). The main approach will be to adjust for baseline variables associated with missing outcome data. We will of course be making strenuous efforts to minimise the amount of missing data, and in our power calculations have estimated that there will be up to 26% missing data at 8 weeks, based on meta-analysis of existing digital treatments for depression trial data (Torous et al., 2020).

Power calculation

The best estimate of a minimum clinically important difference in PHQ-8 is between 11% and 14%, with a standard deviation of 0.32-0.38 (Salaminius et al., 2017; Lewis et al., 2019). In order to provide 80% power at the two sided 5% significance level a total sample size of 378 will be required. Allowing for 26% attrition, we will aim to recruit 238 participants per arm. Power calculations were carried out using Stata 16.

Minimum clinically important difference	Standard deviation	Total sample size	Allowing for 26% attrition
0.11	0.32	268	338
0.14	0.32	166	210
0.11	0.38	378	476
0.14	0.38	234	294

References

Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders*. 2004 Jul 1;81(1):61-6.

Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *Journal of affective disorders*. 2009 Apr 1;114(1-3):163-73.

Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. *Psychiatry investigation*. 2019 Apr;16(4):300.

White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. *Statistics in medicine*. 2011 Nov 30;30(27):3192-207.

Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*. 2010 Dec;11(1):1-8.

Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *Journal of affective disorders*. 2020 Feb 15;263:413-9.

Lewis G, Duffy L, Ades A, Amos R, Araya R, Brabyn S, Button KS, Churchill R, Derrick C, Dowrick C, Gilbody S. The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *The Lancet Psychiatry*. 2019 Nov 1;6(11):903-14.

Katon W, Robinson P, Von Korff M, Lin E, Bush T, Ludman E, Simon G, Walker E. A multifaceted intervention to improve treatment of depression in primary care. *Archives of general psychiatry*. 1996 Oct 1;53(10):924-32.

Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, Sarris J. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*. 2017 Oct;16(3):287-98.

Salaminios G, Duffy L, Ades A, Araya R, Button KS, Churchill R, Croudace T, Derrick C, Dixon P, Dowrick C, Gilbody S. A randomised controlled trial assessing the severity and duration of depressive symptoms associated with a clinically significant response to sertraline versus placebo, in people presenting to primary care with depression (PANDA trial): study protocol for a randomised controlled trial. *Trials*. 2017 Dec;18(1):1-4.