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.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Measure</th>
<th>Baseline</th>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
<th>Week 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>sociodemographics</td>
<td>Bespoke questionnaire</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression symptoms</td>
<td>PHQ-8</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Health related quality of life</td>
<td>SF-12</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>User experience</td>
<td>MAUQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Sleep</td>
<td>Time asleep each night</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep</td>
<td>Time in bed each night</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Daily steps</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart rate variability</td>
<td>Heart rate variability</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menstrual cycle</td>
<td>Period and predicted fertile window</td>
<td>Monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workouts</td>
<td>Daily workout time</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wellbeing</td>
<td>Likert scale</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mood/energy</td>
<td>Circumplex model</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medication adherence</td>
<td>Yes/no</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal</td>
<td>Free text</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>Temperature/ cloud cover/rain</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air pollution</td>
<td>Air quality index</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pollen count</td>
<td>Weeds/trees/grasses</td>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 (Salaminios 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.

<table>
<thead>
<tr>
<th>Minimum clinically important difference</th>
<th>Standard deviation</th>
<th>Total sample size</th>
<th>Allowing for 26% attrition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.11</td>
<td>0.32</td>
<td>268</td>
<td>338</td>
</tr>
<tr>
<td>0.14</td>
<td>0.32</td>
<td>166</td>
<td>210</td>
</tr>
<tr>
<td>0.11</td>
<td>0.38</td>
<td>378</td>
<td>476</td>
</tr>
<tr>
<td>0.14</td>
<td>0.38</td>
<td>234</td>
<td>294</td>
</tr>
</tbody>
</table>

**References**


