

1

2 Digitally managing depression: a fully remote randomized attention-
3 placebo controlled trial

4

5 Aaron Kandola^{1,2}, Kyra Edwards³, Marie AE Muller³, Bettina Dührkoop², Bettina Hein², Joris
6 Straatman², Joseph F Hayes^{*2,3,4}

7

8 1 MRC Unit of Lifelong Health and Aging, UCL

9 2 juli Health

10 3 Division of Psychiatry, UCL

11 4 Camden and Islington NHS Foundation Trust

12

13 *Corresponding author:

14 Joseph Hayes

15 6th Floor Maple House,

16 149 Tottenham Court Road,

17 London

18 joseph.hayes@ucl.ac.uk

19

20 **Abstract**

21 **Background**

22 Depression is a common and disabling condition. Digital apps may augment or facilitate care,
23 particularly in under-served populations. We tested the efficacy of juli, a digital self-management
24 app for depression in a fully remote randomized controlled trial.

25

26 **Methods**

27 We completed a pragmatic single-blind trial of juli for depression. We included participants aged
28 over 18 who self-identified as having depression and scored 5 or more on the Patient Health
29 Questionnaire-8 (PHQ-8). Participants were randomly assigned (1:1) to receive juli for 8 weeks or a
30 limited attention-placebo control version of the app. Our primary outcome was the difference in
31 PHQ-8 scores at 8 weeks. Secondary outcomes were remission, minimal clinically important
32 difference, worsening of depression, and health-related quality of life. Analyses were per protocol
33 (primary) and modified intention-to-treat (secondary). The trial was registered at the ISRCTN registry
34 (ISRCTN12329547).

35

36 **Results**

37 Between May 2021 and January 2023, we randomised 908 participants. 662 completed the week 2
38 outcome assessment and were included in the modified intention-to-treat analysis, and 456
39 completed the week 8 outcome assessments (per protocol). The mean baseline PHQ-8 score was
40 consistent with a diagnosis of moderately severe depression. In the per-protocol analysis, the juli
41 group had a lower mean PHQ-8 score (10.78, standard deviation 6.26) than the control group (11.88,
42 standard deviation 5.73) by week 8 (baseline adjusted β -coefficient -0.94, 95%CI -1.87 to -0.22,
43 $p=0.045$). Remission and minimal clinically important difference were increased in the juli group at 8
44 weeks (adjusted odds ratio 2.22, 95%CI 1.45-3.39, $p<0.001$ and adjusted odds ratio 1.56, 95%CI 1.08
45 to 2.27, $p=0.018$). There were no between-group differences in health-related quality of life physical
46 or mental component scores or worsening of depression.

47

48 **Conclusion**

49 Use of juli reduced symptoms of depression at 8 weeks compared with an attention-placebo control.
50 The juli app is a digital self-management tool that could increase accessibility of evidence-based
51 depression treatments.

52

53 Introduction

54 Depression is a major contributor to the global burden of disease. Each year, millions of people are
55 diagnosed with depression, and is likely to be the leading cause of disability in high-income countries
56 by 2030¹. The mainstay of treatment for depression includes psychological intervention and
57 antidepressant medication. Barriers to accessing care include high costs, low availability, long
58 waiting lists and stigma. Digital interventions may remove some of these barriers by offering scalable
59 solutions that are convenient and timely for people with depression².

60
61 Recent meta-analyses of randomised controlled trials (RCTs) of smartphone app-based psychological
62 interventions for depression symptoms find a small to moderate reduction in symptoms compared
63 to placebo, such as waiting list control^{3,4}. However, subgroup analyses find no difference in
64 depressive symptoms in studies with active control groups, which is corroborated by the findings of
65 other systematic reviews⁵. Reviews also only include small numbers of trials, which are frequently of
66 short duration and/or include few participants. Participants often do not have severe depression³
67 Many of the apps included in the reviews are no longer (or have never been) commercially available,
68 highlighting that translation from research to clinical impact is difficult in this space. Other
69 challenges include low levels of retention and engagement with digital health apps in trials and real-
70 world use⁶. The reviews all conclude that further methodologically robust RCTs are required. We
71 sought to address some of the limitations of previous RCTs by pre-registering our trial, powering it
72 adequately to detect between-group differences, following-up users for an adequate duration,
73 efforts to minimise attrition and avoiding recruitment from unrepresentative populations (such as
74 from mood disorders clinics). Our RCT was fully remote, increasing cost-effectiveness, time efficiency
75 and reach.

76
77 The digital health app juli, aims to support people with depression via numerous evidence-based
78 approaches, including mood tracking, medication reminders, positive affect journaling, data
79 visualisation of sleep, activity, exercise, heart rate variability and behavioural activation
80 recommendations about how to improve these parameters⁷. As such, it combines many of the
81 elements that have been found to be effective in research-grade apps for depression. However,
82 there has been less evaluation of consumer-grade apps in real-life practice. This is important as
83 many popular health apps have not been scrutinised in the way new interventions traditionally
84 would be. We hypothesised that individuals randomized to juli would have a greater reduction in
85 depression symptoms at 8 weeks than those in an attention placebo control group.

87 Methods

88 Study design and participants

89 We conducted a fully remote pragmatic single-blind, placebo control randomised controlled trial of
90 people with depression from anywhere in the world. Individuals were eligible for inclusion if they were
91 aged 18 to 65, were English speaking, had access to an iPhone, and self-identified as having
92 depression, with a score of 5 or more on the Patient Health Questionnaire 8-item version (PHQ-8) at
93 baseline. We recruited participants via online adverts, social media posts and self-help groups for
94 depression. Recruitment was from May 2021 until January 2023. All participants provided written
95 informed consent via a consent form within the app. Ethical approval was from the University
96 College London Ethics Committee (ID number 19413/001). The trial was registered on the ISRCTN
97 registry (<https://doi.org/10.1186/ISRCTN12329547>).

99 Randomisation and masking

100 We randomly assigned participants (1:1) to a full version of juli or an attention-placebo control
101 version. Block randomisation was conducted within the app using automated code, with random
102 block sizes between 4 and 8. The researchers and independent statisticians were masked to
103 treatment allocation until the completion of the analysis.

104

105 **Procedures**

106 Participants were either allocated to the full version of the juli app or an attention placebo control
107 app group. If allocated to the full version of the app, participants were prompted to open the app
108 each day via an automated alert at the time of their choosing. They were asked to rate how they
109 were feeling on a scale using 5 emoji faces and a circumplex model with mood on the x-axis and
110 energy on the y-axis⁸. Individuals were also able to track things that they considered as important
111 contributors to their mood⁹. The app passively gathered information via smartphone and
112 smartwatch sensors on sleep, activity, workouts, menstrual cycle, and heart rate variability on a daily
113 basis and presented this data to the participant, showing associations with mood¹⁰. The app then
114 provided recommendations about these parameters to guide healthy behaviours via behavioural
115 activation¹¹. The app includes a medication reminder function that can be set by the participants to
116 improve medication adherence¹². Participants were also encouraged to engage in positive affect
117 journaling via the app (Figure 1)¹³. The juli app was designed by a psychiatrist and experts in
118 gamification¹⁴. It is grounded in evidence-based treatments. Participants were guided towards all
119 elements of the app, but did not have to engage with elements they did not like. If allocated to the
120 control arm, participants were prompted to open the app each day via an automated alert and to
121 rate how they were feeling on a scale using 5 emoji faces. Control participants did not receive any
122 further intervention.

123

124 Baseline assessments and follow-up assessments at two, four, six and eight weeks were all
125 completed within the app. Participants completed the PHQ-8 for depression symptoms and the 12-
126 Item Short Form Health Survey (SF-12) for health-related quality of life at baseline. The PHQ is a
127 widely used self-completed depression scale that is recommended by the Common Measures in
128 Mental Health Science Governance board as one of a core list of research questionnaires that should
129 be used by funded researchers. The PHQ-9 closely matches the DSM-IV criteria for a major
130 depressive episode and may be more sensitive to change than other measures of depression, such
131 as the Hamilton Rating Scale for Depression and the Beck Depression Inventory¹⁵. The PHQ-8
132 excludes the question about suicidality, which is preferred in studies where patient contact is
133 remote, such as via digital technologies or telephone. Research indicates that the deletion of this
134 question has little effect on the scale's psychometric properties because this question is the least
135 frequently endorsed item on the PHQ-9. Subsequently, the PHQ-8 has identical scoring thresholds
136 for depression severity, with higher scores representing more severe depression^{16,17}. The SF-12 is a
137 self-reported health status measure¹⁸. Possible scores range from 0 to 100, with higher scores
138 indicating better quality of life.

139

140 **Outcomes**

141 The primary outcome was the total score on the PHQ-8 at 8 weeks.

142

143 The secondary outcomes were: 1) PHQ-8 score as a continuous outcome at 2, 4, 6 and 8 weeks in a
144 repeated measures analysis, 2) PHQ-8 score as a binary outcome where remission is a score of <10
145 at 8 weeks, 3) remission at 2, 4, 6 and 8 weeks in a repeated measures analysis, 4) Difference in SF-
146 12 physical and mental component scores at 8 weeks and, 5) SF-12 physical and mental component
147 scores at 4 and 8 weeks in a repeated measures analysis.

148

149 We added post-hoc outcomes that included: 1) achieving a minimal clinically important difference
150 (MCID) at 8 weeks defined by the effective dose 50 method, which accounts for baseline severity
151 and is the smallest difference in PHQ-8 scores that are of perceived benefit¹⁹, 2) a worsening of
152 depression, defined as a >20% increase in PHQ-8 from baseline.

153

154 **Statistical analyses**

155 Our analysis plan was pre-printed (<https://discovery.ucl.ac.uk/id/eprint/10129350/>) and included on
156 the ISRCTN registry. We followed the Consolidated Standards of Reporting Trials (CONSORT)
157 guidelines in reporting and analysing our data¹⁷. Primary and secondary outcomes were described in
158 the published protocol and on the ISRCTN registry before the study started.

159
160 The best estimate of a MCID in PHQ-8 is between 11% and 14%, with a standard deviation of 0.32-
161 0.38^{20,21}. 80% power at the two-sided 5% significance level requires a total sample size of 378.
162 Allowing for 26% attrition²², we aimed to recruit 238 participants per arm. Power calculations were
163 carried out using Stata.

164
165 The primary outcome was the difference in total PHQ-8 score at 8 weeks between control and
166 intervention groups in a per protocol analysis. This was estimated with a linear regression model
167 adjusted for baseline PHQ-8. We calculated the odds ratio of remission at 8 weeks (PHQ-8<10),
168 achieving MCID and worsening of depression, adjusting for baseline severity using logistic
169 regression. Repeat measures analyses were completed using linear or logistic mixed effect models
170 adjusting for baseline severity.

171
172 We also examined all outcomes in a modified intention-to-treat analysis including all randomised
173 participants with a complete baseline and week 2 PHQ-8. Missing outcome data were imputed by
174 last observation carried forward and controlled multiple imputation²³.

175
176 All analyses were completed by independent statisticians using Stata and R, who have no financial
177 conflict of interest with the company providing the juli app.

178 179 **Results**

180 **Baseline characteristics**

181 We recruited 456 individuals who were retained in the trial for 8 weeks and formed the basis of our
182 primary per protocol analysis (Figure 2). The majority of participants were female and had
183 experienced depression for more than five years (Table 1). The majority were diagnosed by a
184 physician and continued to be in regular or occasional contact with a doctor about their depression.
185 The mean PHQ-8 score at baseline was 16.16 (standard deviation (SD) 4.71).

186 187 **Per protocol analysis**

188 At 8 weeks, participants in the intervention group had a mean PHQ-8 score of 10.78 (SD 6.26) and
189 participants in the control group had a mean of 11.88 (SD 5.73). After accounting for baseline PHQ-8
190 score the intervention group had lower depression symptom scores at 8 weeks (-0.94, 95%
191 confidence interval (CI) -1.87 to -0.22, p=0.045). The odds of being in remission by week 8 was
192 higher in the intervention group after accounting for baseline depression severity (adjusted odds
193 ratio 2.22, 95%CI 1.45 to 3.39, p<0.001). Participants in the intervention group were more likely to
194 experience MCID (adjusted odds ratio 1.56, 95%CI 1.08 to 2.27, p=0.018) than those in the control
195 group. Repeat measures analyses of these outcomes at 2, 4, 6 and 8 weeks suggest that this effect
196 was maintained over time (Supplement Table 2). We found no effect of the intervention on SF-12
197 mental or physical component scores. The odds of experiencing worsening symptoms were similar in
198 the intervention and control groups (adjusted odds ratio 0.83, 95%CI 0.38 to 1.81, p=0.633).

199 200 **Intention-to-treat analysis**

201 Modified intention-to-treat analysis included 322 participants in the intervention group and 340
202 controls who completed baseline and week 2 PHQ-8 (Supplement Table 1). Baseline characteristics
203 of participants in these groups were similar to the per protocol analysis. In the last observation
204 carried forward dataset, there was no clear difference between intervention and control groups
205 after accounting for baseline severity (β -coefficient -0.72, 95%CI -1.49 to 0.06)(Supplement table 2)

206 at week 8. However, the repeated measures analysis of PHQ-8 scores found a positive effect of the
207 intervention on depression symptoms. The odds ratios for remission at 8 weeks and remission in
208 repeated measures were similar to the per protocol analysis, and suggested higher odds of remission
209 in the intervention group. The odds ratios for MCID and worsening of symptoms were consistent
210 with the per protocol results. Results from the multiple imputation dataset were consistent with the
211 last observation carried forward findings.

212

213 Discussion

214 We found a small reduction in depression symptoms in participants using juli for 8 weeks compared
215 to an attention-placebo control. Both groups had a reduction in PHQ-8 scores over the course the
216 trial. The mean intervention arm PHQ-8 score decreased by 37%, a reduction considerably greater
217 than the MCID²⁴. Participants allocated to juli were more than twice as likely to be in remission by 8
218 weeks, and more likely meet the threshold for MCID in depression symptoms.

219

220 The participants had a mean baseline PHQ-8 score consistent with a diagnosis of moderately severe
221 depression, the majority had longstanding depression and were under the ongoing care of a
222 physician. Therefore these participants were chronically unwell and potentially experiencing
223 depressive symptoms that were difficult to treat. This is important as our participants differ from
224 those included in many digital depression intervention RCTs in terms of severity and duration³.

225

226 The improvement in the attention-control group was of a similar magnitude to other placebo
227 controlled trials of depression interventions²⁵. In addition, even basic mood monitoring, as required
228 by our control group, has been found to decrease depression symptoms²⁶. This may have reduced
229 difference observed between intervention and control groups, compared to an inactive control (such
230 as waiting list).

231

232 Findings from meta-analyses support the use of digital technologies for the treatment of depression.
233 However, many of the interventions reviewed are not available to patients, often because they are
234 not available commercially or via healthcare providers. juli uses a combination of evidence-based
235 approaches to support symptom reduction in depression and is available globally on Apple and
236 Android formats. Our trial provides methodologically robust evidence of efficacy.

237

238 Strengths and limitations

239 Our RCT has a number of strengths and limitations. We successfully recruited, screened,
240 randomized, treated, and assessed a geographically dispersed sample of participants. We modified
241 the juli app so that for RCT participants could consent, be randomized and take the baseline
242 assessment within the app. This facilitated global recruitment, with low cost, in a pragmatic manner,
243 with good external validity. However, this also meant that we lacked information on potentially
244 important baseline characteristics, such as social determinants of health, as we did not want to
245 overburden the participants. Despite this, balance in the recorded baseline characteristics after
246 randomisation supports the assumption that randomisation was successful. Additionally, we did not
247 include a large battery of outcome measures, which may have shed further light on our findings. For
248 example, improvements in anxiety symptoms have been found in RCTs of interventions for
249 depression²¹.

250

251 We pre-registered our analysis plan and did not deviate from this. Attrition was higher than we
252 anticipated (49.78% from randomisation to week 8). The majority of the attrition occurred between
253 randomisation and week 2, which is common in RCTs, including for depression apps. A recent meta-
254 analysis of dropout rates found similar attrition after adjusting for publication bias²². We overcame
255 this by recruiting until we had sufficient numbers who had completed the week 8 outcome measures
256 and examined differences in completers vs non-completers.

257

258 We analysed participants using the app for 8 weeks in our primary analysis to focus on the effects of
259 maintained use. Baseline characteristics were similar in the modified intention-to-treat and primary
260 analysis, and results were consistent, suggesting little difference between completers and non-
261 completers. Following the publication of our protocol and commencement of our RCT, newer
262 research on the PHQ-8 MCID was published suggesting that we may have underestimated our power
263 calculations^{19,24}. We used these newer methods to derive a post hoc MCID outcome¹⁹. We used two
264 imputation methods for the intention-to-treat analysis that make different assumptions²³. Results
265 from both methods did not differ.

266

267 We found that most participants were diagnosed by a doctor and remained under their care. This
268 highlights the severity of our participants depression, but does not suggest that people were
269 accessing juli because of an absence of traditional care. Individuals with no previous or ongoing
270 healthcare may differentially benefit from digital technologies. Digital apps could be one solution to
271 an overburdened healthcare system, particularly in groups or areas where accessing treatment for
272 mental health conditions is challenging or stigmatising²⁷. However, some people may still be
273 unwilling or unable to use apps for health, despite their relative ease of access. The majority of
274 participants were female, which reflects established differences in sex-specific rates of depression
275 and help-seeking for depression²⁸. Some of our participants (~7%) identified as transgender or non-
276 binary. This group is under-represented in research but has higher risk of depression and other
277 mental disorders²⁹. Our high uptake suggests that digital technologies may be a better way of
278 engaging these populations.

279

280 An even bigger problem than the high attrition in RCTs of digital apps for depression is their lack of
281 real-world retention and engagement. The proportion of 30-day retention is less than 10% across
282 mental health apps⁶. The 30-day retention of juli users (non-RCT participants) is 25%. This highlights
283 that engagement and the potential clinical benefit can be increased by methods employed by juli.
284 However, it is unclear which specific features increase engagement. For example, a 2021 meta-
285 analysis found no benefit of gamification³⁰. More research is needed to understand which
286 depression app features are integral to improving mental health symptoms.

287

288 **Conclusion**

289 The juli app can reduce depressive symptoms within 8 weeks, with increased probability of
290 remission and MCID. As such it represents a low-risk addition to the care package of people with
291 mild to severe depression. Further research is required to determine the most cost-effective
292 technical support processes to enhance engagement, and how juli could be implemented in current
293 public health or clinical care models.

294

295 **Author Contributions**

296 JFH conceived the study; JFH, AK, BD, BH and JS designed the study; AK, BD and BH collected the
297 data; KE and MM analysed the data; JFH wrote the initial draft; all authors edited and approved the
298 final manuscript.

299

300 **Acknowledgements**

301 AK is supported by the UK Research and Innovation (UKRI) Digital Youth Programme award [MRC
302 project reference MR/W002450/1], which is part of the AHRC/ESRC/MRC Adolescence, Mental
303 Health and the Developing Mind programme. JFH is supported by the UK Research and Innovation
304 grant MR/V023373/1, the University College London Hospitals NIHR Biomedical Research Centre,
305 and the NIHR North Thames Applied Research Collaboration.

306

307 **Conflicts of Interest**

308 The current study was funded by juli Health. AK, BD, BH, JS and JFH are shareholders in juli Health.
309 AK has received consultancy fees from juli Health and Wellcome Trust. BD, BH, JS and JFH are a co-
310 founders of juli Health. JFH has received consultancy fees from juli Health and Wellcome Trust. KE
311 and MM have no conflicts of interest. The funders played no part in the analysis of the data.

312

313 **Data Availability**

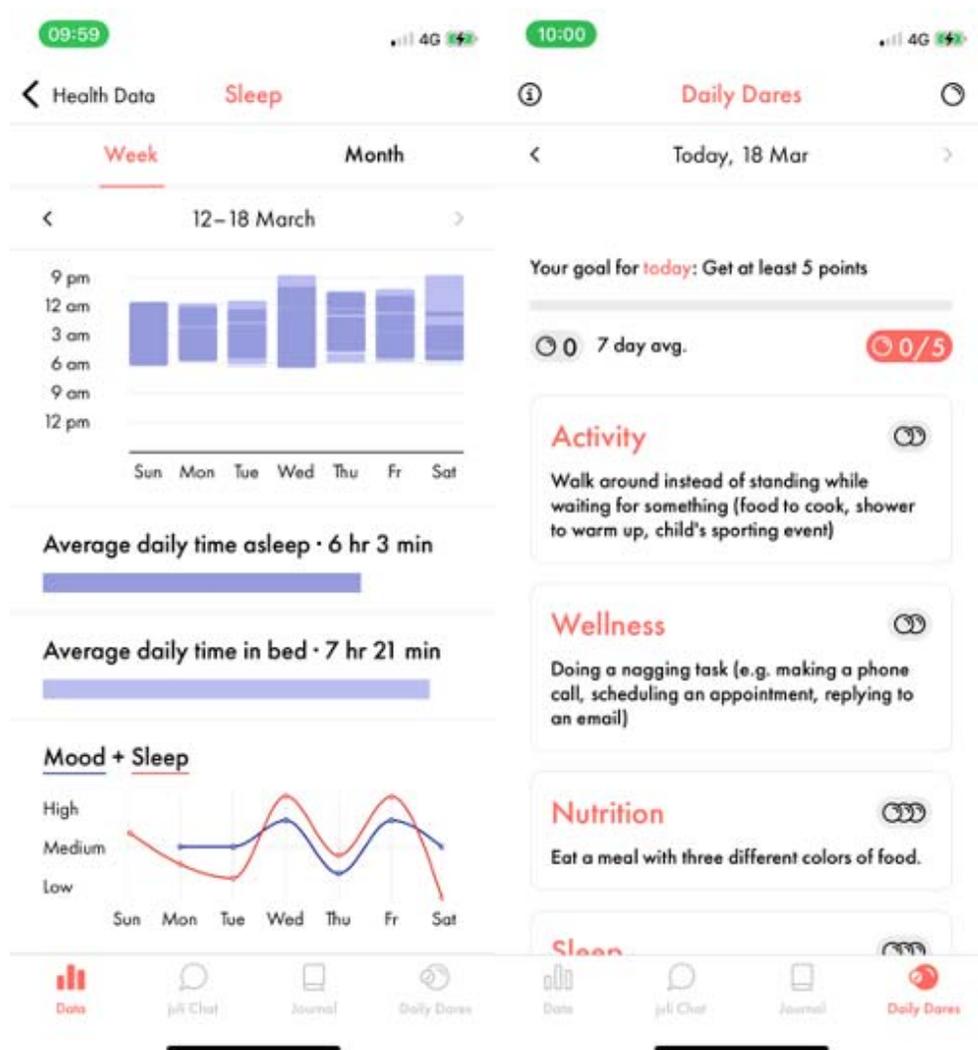
314 All data produced in the present study are available upon reasonable request to the authors.

315

316

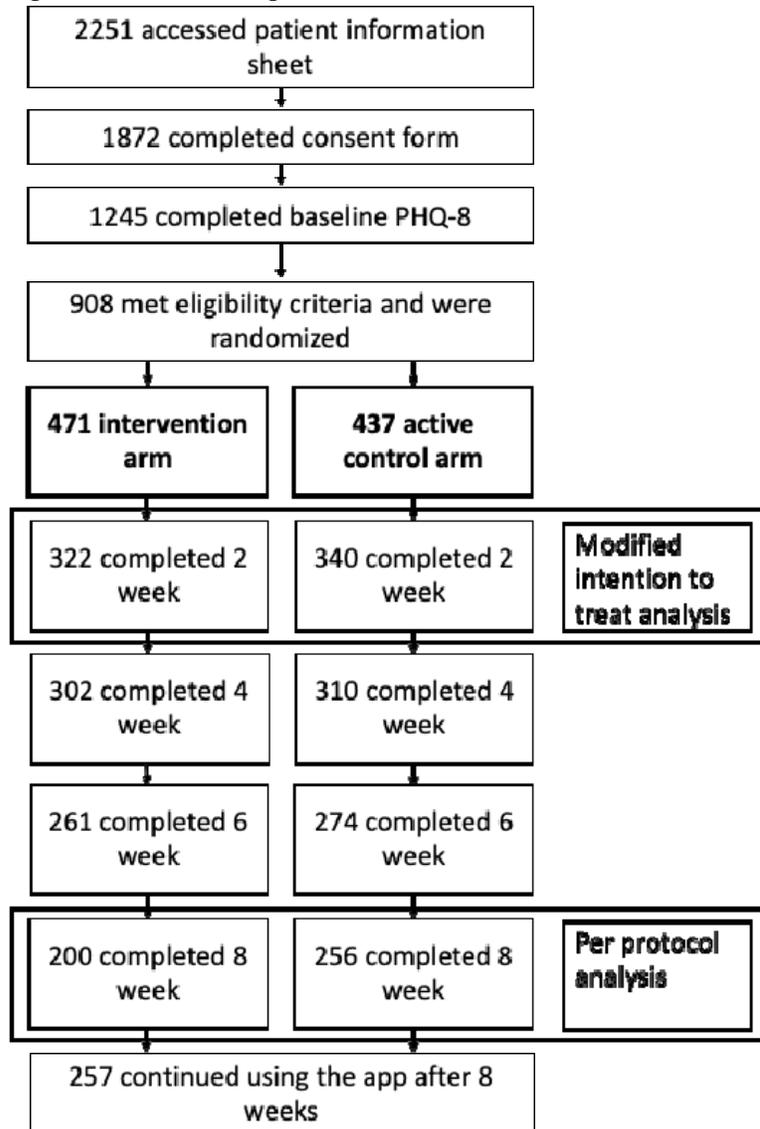
317
318
319

Figure 1. Example juli screen shots



320
321
322

323 **Figure 2. CONSORT diagram**



324
325
326

327 **Table 1. Baseline characteristics**
328

	Intervention (n=200)	Control (n=256)	All (n=456)
Age	34.39 (12.41)	32.65 (12.28)	33.41 (12.35)
Gender			
Female	154 (77.00)	202 (78.91)	356 (78.07)
Male	33 (16.50)	37 (14.45)	72 (15.69)
Other	13 (6.50)	17 (6.64)	30 (6.58)
Depression duration			
<1 month	1 (0.50)	1 (0.39)	2 (0.44)
1 to <3 months	7 (3.50)	2 (0.78)	9 (1.97)
3 months to <1 year	2 (1.00)	14 (5.47)	16 (3.51)
1 year to <2 years	11 (5.50)	15 (5.86)	26 (5.70)
2 years to <5 years	43 (21.50)	49 (19.14)	92 (20.18)
>5 years	136 (68.20)	175 (68.36)	311 (68.20)
Physician contact			
Regular	76 (38.00)	112 (43.75)	188 (41.23)
Occasional	67 (33.50)	64 (25.00)	131 (28.73)
Not anymore	32 (16.00)	38 (14.84)	70 (15.35)
Never	25 (12.50)	42 (16.41)	67 (14.69)
Diagnosed by a physician			
Yes	179 (88.61)	219 (85.55)	397 (87.06)
No	22 (11.00)	37 (14.45)	59 (12.94)
PHQ-8 total score ¹	16.09 (4.94)	16.30 (4.67)	16.21 (4.78)
SF-12 physical health subscale ²	46.61 (9.29)	45.05 (9.81)	45.74 (9.61)
SF-12 mental health subscale ³	22.15 (7.87)	22.09 (7.70)	22.12 (7.79)

329 Data are n (%) or mean (SD). ¹Patient Health Questionnaire, 8-item version (possible range 0–24),
330 ²Short-Form Health Survey12 physical health subscale (possible range 0–100), ³Short-Form Health
331 Survey-12 mental health subscale (possible range 0–100). Data used in per protocol analysis of
332 individuals completing week 8 PHQ-8
333

334

335 **References**

336

- 337 1 Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease
338 from 2002 to 2030. *PLoS medicine* **3**, e442 (2006).
- 339 2 Moshe, I. *et al.* Digital interventions for the treatment of depression: A meta-analytic
340 review. *Psychological bulletin* **147**, 749 (2021).
- 341 3 Serrano-Ripoll, M. J., Zamanillo-Campos, R., Fiol-DeRoque, M. A., Castro, A. & Ricci-
342 Cabello, I. Impact of Smartphone App-Based Psychological Interventions for
343 Reducing Depressive Symptoms in People With Depression: Systematic Literature
344 Review and Meta-analysis of Randomized Controlled Trials. *JMIR mHealth and*
345 *uHealth* **10**, e29621 (2022).
- 346 4 Astafeva, D. *et al.* The efficacy of mobile phone-based interventions for the
347 treatment of depression: A systematic meta-review of meta-analyses of randomized
348 controlled trials. *Psychiatria Danubina* **34**, 155-163 (2022).
- 349 5 Hernández-Gómez, A., Valdés-Flrido, M. J., Lahera, G. & Andrade-González, N.
350 Efficacy of smartphone apps in patients with depressive disorders: A systematic
351 review. *Frontiers in Psychiatry*, 1166 (2022).
- 352 6 Baumel, A., Muench, F., Edan, S. & Kane, J. M. Objective user engagement with
353 mental health apps: systematic search and panel-based usage analysis. *Journal of*
354 *medical Internet research* **21**, e14567 (2019).
- 355 7 Firth, J. *et al.* The efficacy of smartphone-based mental health interventions for
356 depressive symptoms: a meta-analysis of randomized controlled trials. *World*
357 *Psychiatry* **16**, 287-298 (2017).
- 358 8 Posner, J., Russell, J. A. & Peterson, B. S. The circumplex model of affect: An
359 integrative approach to affective neuroscience, cognitive development, and
360 psychopathology. *Development and psychopathology* **17**, 715-734 (2005).
- 361 9 Dubad, M., Winsper, C., Meyer, C., Livanou, M. & Marwaha, S. A systematic review
362 of the psychometric properties, usability and clinical impacts of mobile mood-
363 monitoring applications in young people. *Psychological medicine* **48**, 208-228 (2018).
- 364 10 Colombo, D. *et al.* Current state and future directions of technology-based ecological
365 momentary assessment and intervention for major depressive disorder: a systematic
366 review. *Journal of clinical medicine* **8**, 465 (2019).
- 367 11 Ekers, D. *et al.* Behavioural activation for depression; an update of meta-analysis of
368 effectiveness and sub group analysis. *PLoS one* **9**, e100100 (2014).
- 369 12 Hammonds, T. *et al.* Adherence to antidepressant medications: a randomized
370 controlled trial of medication reminding in college students. *Journal of American*
371 *College Health* **63**, 204-208 (2015).
- 372 13 Smyth, J. M. *et al.* Online positive affect journaling in the improvement of mental
373 distress and well-being in general medical patients with elevated anxiety symptoms:
374 A preliminary randomized controlled trial. *JMIR mental health* **5**, e11290 (2018).
- 375 14 Wu, A. *et al.* Smartphone apps for depression and anxiety: a systematic review and
376 meta-analysis of techniques to increase engagement. *NPJ digital medicine* **4**, 20
377 (2021).
- 378 15 Löwe, B., Kroenke, K., Herzog, W. & Gräfe, K. Measuring depression outcome with a
379 brief self-report instrument: sensitivity to change of the Patient Health
380 Questionnaire (PHQ-9). *Journal of affective disorders* **81**, 61-66 (2004).

- 381 16 Kroenke, K. *et al.* The PHQ-8 as a measure of current depression in the general
382 population. *Journal of affective disorders* **114**, 163-173 (2009).
- 383 17 Shin, C., Lee, S.-H., Han, K.-M., Yoon, H.-K. & Han, C. Comparison of the usefulness of
384 the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of
385 psychiatric outpatient data. *Psychiatry investigation* **16**, 300 (2019).
- 386 18 Jenkinson, C. *et al.* A shorter form health survey: can the SF-12 replicate results from
387 the SF-36 in longitudinal studies? *Journal of Public Health* **19**, 179-186 (1997).
- 388 19 Bauer-Staeb, C. *et al.* Effective dose 50 method as the minimal clinically important
389 difference: Evidence from depression trials. *Journal of Clinical Epidemiology* **137**,
390 200-208 (2021).
- 391 20 Salaminios, G. *et al.* A randomised controlled trial assessing the severity and duration
392 of depressive symptoms associated with a clinically significant response to sertraline
393 versus placebo, in people presenting to primary care with depression (PANDA trial):
394 study protocol for a randomised controlled trial. *Trials* **18**, 1-14 (2017).
- 395 21 Lewis, G. *et al.* The clinical effectiveness of sertraline in primary care and the role of
396 depression severity and duration (PANDA): a pragmatic, double-blind, placebo-
397 controlled randomised trial. *The Lancet Psychiatry* **6**, 903-914 (2019).
- 398 22 Torous, J., Lipschitz, J., Ng, M. & Firth, J. Dropout rates in clinical trials of smartphone
399 apps for depressive symptoms: a systematic review and meta-analysis. *Journal of*
400 *affective disorders* **263**, 413-419 (2020).
- 401 23 Cro, S., Morris, T. P., Kenward, M. G. & Carpenter, J. R. Sensitivity analysis for clinical
402 trials with missing continuous outcome data using controlled multiple imputation: a
403 practical guide. *Statistics in medicine* **39**, 2815-2842 (2020).
- 404 24 Kounali, D. *et al.* How much change is enough? Evidence from a longitudinal study on
405 depression in UK primary care. *Psychological Medicine* **52**, 1875-1882 (2022).
- 406 25 Jones, B. D. *et al.* Magnitude of the placebo response across treatment modalities
407 used for treatment-resistant depression in adults: a systematic review and meta-
408 analysis. *JAMA Network Open* **4**, e2125531-e2125531 (2021).
- 409 26 Kauer, S. D. *et al.* Self-monitoring using mobile phones in the early stages of
410 adolescent depression: randomized controlled trial. *Journal of medical Internet*
411 *research* **14**, e1858 (2012).
- 412 27 Bucci, S., Schwannauer, M. & Berry, N. The digital revolution and its impact on
413 mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*
414 **92**, 277-297 (2019).
- 415 28 Paykel, E. S. Depression in women. *The British Journal of Psychiatry* **158**, 22-29
416 (1991).
- 417 29 Newcomb, M. E. *et al.* High burden of mental health problems, substance use,
418 violence, and related psychosocial factors in transgender, non-binary, and gender
419 diverse youth and young adults. *Archives of sexual behavior* **49**, 645-659 (2020).
- 420 30 Six, S. G., Byrne, K. A., Tibbett, T. P. & Pericot-Valverde, I. Examining the
421 effectiveness of gamification in mental health apps for depression: systematic
422 review and meta-analysis. *JMIR mental health* **8**, e32199 (2021).
- 423