Journal of Affective Disorders 286 (2021) 228–238

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

A therapist-guided smartphone app for major depression in young adults: A randomized clinical trial

Anu Raevuori [a,b,*], Tero Vahlberg [c], Tellervo Korhonen [d], Outi Hilgert [e], Raija Aittakumpu-Hyden [f], Valerie Forman-Hoffman [e]

[a] Department of Adolescent Psychiatry, Helsinki University Hospital, Helsinki, Finland
[b] Clinicum, Department of Public Health, University of Helsinki, Helsinki, Finland
[c] Department of Biostatistics, University of Turku, Turku, Finland
[d] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland
[e] Meru Health Inc. Palo Alto, The United States & Helsinki, Finland
[f] Finnish Student Health Service, Jyvaskyla, Finland

ARTICLE INFO

Keywords:
Depression
Depressive disorder
Randomized controlled trial
Smartphone
Internet-based intervention

ABSTRACT

Background: Meru Health Program (MHP) is a therapist-guided, 8-week intervention for depression delivered via smartphone. The aim was to test its efficacy in patients with clinical depression in a Finnish university student health service.

Methods: Patients (n=124, women 72.6%, mean age 25y) were stratified based on antidepressant status, and randomized into intervention group receiving MHP plus treatment as usual (TAU), and control group receiving TAU only. Depression, measured by the Patient Health Questionnaire-9 (PHQ-9) scale, was the primary outcome. After baseline (T0), follow-ups were at mid-intervention (T4), immediately post-intervention (T8); 3 months (T20), and 6 months (T32) post-intervention.

Results: The intervention group and control group did not have significant differences in depression outcomes throughout end of treatment and follow-up. Among secondary outcomes, increase in resilience (d=0.32, p=0.03) and mindfulness (d=0.57, p=0.002), and reduction in perceived stress (d=-0.52, p=0.008) were greater in MHP+TAU versus TAU at T32; no differences were found in anxiety, sleep disturbances, and quality of life between groups. Post-hoc comparisons of patients on antidepressants showed significantly greater reduction in depression at T32 for MHP+TAU versus TAU (d=-0.73, p=0.01); patients not on antidepressants showed no between-group differences.

Limitations: Limitations include unknown characteristics of TAU, potential bias from patients and providers not being blinded to treatment group, and failure to specify examination of differences by antidepressant status in the protocol.

Conclusions: Most outcomes, including depression, did not significantly differ between MHP+TAU and TAU. Exploratory analysis revealed intervention effect at the end of the 6-month follow-up among patients on antidepressant medication.

1. Introduction

The rate of untreated depression is estimated to exceed 50% in developed countries (Kohn et al., 2004), and its disease burden continues rising globally (Malhi and Mann, 2018). Traditional ways of providing mental health care were hampered by insufficient availability and low scaling (Barkil-Oteo, 2013) already before the outburst of a COVID-19 pandemia, which has mounted further challenges to the field by increasing the need of psychological support (World Health Organization, 2020) and simultaneously demanding physical distancing. Digitally

Abbreviations: AD, antidepressant medication; CBT, cognitive-behavioral therapy; LOCF, last observation carried forward; LSM, least squares means; MHP, Meru health program; FSHS, finnish student health service; RCT, randomized clinical trial; TAU, treatment as usual; T0, baseline; T4, mid-intervention follow-up; T8, post-intervention follow-up; T20, 3 months post-intervention follow-up; T32, 6 months post-intervention follow-up.

* Corresponding author at: Adolescent Psychiatry Rehabilitation Unit, Ratapihantie 11, P.O. Box 356, Helsinki University Hospital 00029, Finland.
E-mail address: anu.raevuori@helsinki.fi (A. Raevuori).

https://doi.org/10.1016/j.jad.2021.02.007
Received 2 November 2020; Received in revised form 14 January 2021; Accepted 1 February 2021
Available online 5 February 2021

delivered mental health care has potential to address these challenges.

Internet- and smartphone-based programs have demonstrated clinical efficacy in the treatment of depression (Firth et al., 2017; Graham et al., 2020; Kerst et al., 2020) with similar effect sizes and attrition rates as traditional interventions (Carlbring et al., 2018). Predictably, comparisons to inactive controls have yielded larger, moderately positive effects, whereas comparisons to active controls have yielded smaller positive effects (Firth et al., 2017). Mixed results have been reported of the effects of the programs with live-therapist support compared to automated programs; either that the former is larger (Richards and Richardson, 2012), or smaller (Firth et al., 2017). Health care professionals tend to view digital interventions more suitable for mild to moderate than severe depression (Kerst et al. 2020), and evidence from the meta-analysis supports the notion (Firth et al., 2017). However, no difference in the treatment effect sizes between community and primary-/secondary health care settings, where severity of depression is likely to differ, has also been reported in a meta-analysis (Richards and Richardson, 2012). The attitudes towards digital applications for depression treatment have been reported positive both among users and health care professionals (Kerst et al., 2020), which promotes their dissemination.

Delivered via smartphone app, Meru Health Program (MHP) provides a comprehensive intervention by combining elements from three evidence-based interventions for depression - behavioral activation, mindfulness, and cognitive-behavioral therapy (CBT) (Economides et al., 2019). The content includes support by a licensed remote therapist; and text, videos, audio-guided mindfulness meditation exercises, infographics illustrating CBT principles, and journal prompts. In this study, we tested the efficacy of 8-week MHP in a randomized-clinical trial (RCT) among patients with major depression from Finnish Student Health Service (FSHS). Participants were young adults representing a cohort where lifetime prevalence of depression is suggested to be higher than in older cohorts, and an age group by which three-fourths of lifetime cases of mood disorders have had their onset (Richards, 2011).

The primary aim of the study was to compare the efficacy of the MHP plus treatment as usual (TAU) (intervention group) to TAU only (control group) in the change of depression symptoms. Secondary aims were to 1) examine whether the treatment response between the intervention and the control groups differed in anxiety symptoms, sleep problems, perceived stress, quality of life, resilience, mindfulness, and attrition; and 2) explore whether there was an association between the quantity of mindfulness practice and change in depression and anxiety symptoms. We hypothesized that depression symptoms would decrease among patients in the intervention group across the 8-week MHP, and that the difference in the depression score between the intervention and the control group would be statistically and clinically significant immediately post-intervention, remaining at a lower level through the follow-up. We also hypothesized that anxiety, sleep problems and perceived stress would show greater reduction among the intervention group across the study period than those in the control group, remaining at a lower level during the follow-up. We assumed that quality of life, resilience and mindfulness would show greater increase among the intervention group during the intervention, remaining at a higher level at the follow-up. We also assumed that the attrition would be smaller in the intervention group versus control group. Finally, we hypothesized that there would be a dose-response relationship between the level of the engagement to the intervention mindfulness practice regimen and the reduction of depression and anxiety symptoms, and that this would be modified by the internalization of mindfulness skills.

2. Methods

2.1. Study design

We used a randomized-controlled multi-center intervention study design with simple random sampling. Randomization was performed using block randomization by the Clinpal software, a data secure online patient data collection and storage platform for clinical trials by eClinicalHealth Ltd (https://www.clinpal.com/). The random allocation sequence was automatically generated by the Clinpal. Randomization was done in permuted blocks stratified by participants' antidepressant medication (AD) status (self-reported AD status at baseline, yes *versus* no) with a 1:1 allocation ratio using random block sizes of 2, 4, 6, 8 and 10. Researchers, assistants and patients were blind to the stratification and randomization sequence. Statistical analyses were conducted by an independent statistician.

Participants from both groups received TAU, whereas those in the intervention group additionally received 8-week MHP. Baseline (T0) data were gathered within 2-week time frame before the MHP for the intervention groups started. Follow-up points were at mid-intervention (T4); at the intervention completion (T8); 3 months (T20), and 6 months after the intervention completion (T32).

The trial is registered with ISRCTN registry (ISRCTN17156687, accessible at https://www.isrctn.com/ISRCTN17156687. Approvals for the study were provided by the Ethics Committee of the Tampere University Central Hospital (ref: 09/2017, ETL-code R17128) and FSHS (12/2017). Patient safety was monitored by systematic monitoring of adverse/ serious adverse events by therapists and a psychiatrist of the team.

2.2. Participants and procedure

Participants were university students recruited from general practitioners' (GPs, the vast majority) and psychiatrists' appointments in FSHS clinics in 11 cities. FSHS provides nationwide primary level health care for all university students in Finland (coverage n=127 000 students). The access to services is easy and low cost.

GPs and psychiatrists in FSHS were instructed to provide initial information and documents for the study screening for their patients with a diagnosis of major depressive disorder. Diagnosis was required to be made by GP or psychiatrist. If a patient was willing to enroll into the study screening, she/he returned the informed consent for the screening in a pre-paid envelope to MHP therapist, after which a therapist scheduled a screening phone call. In the screening call, therapist provided information of the study, underwent the inclusion and exclusion criteria, and answered patient's questions. After the call, each patient decided whether she/he wanted to participate in the study, and the therapist considered patient's eligibility. Therapists sent an email invitation from the Clinpal contacting eligible patients who could log in and setup an account into the online patient data collection platform, where all self-reported instruments were administered. In the first login to the platform, each patient signed the informed consent for the study, provided details of her/his potential AD, and filled in the baseline questionnaires. These baseline data were gathered before the stratification and randomization, simultaneously within the first login to get information of the AD status for the stratification. Recruitment started in April 10th 2018 and completed June 30th 2019. The follow-up completed March 31st 2020. The intervention was free of charge for patients; no compensation was provided.

2.3. Inclusion and exclusion criteria

Study eligibility criteria included having ICD-10 diagnosis of a major depressive disorder (single or recurrent episode; ICD-10 codes F32.0, F32.1, F32.2, or F33.0, F33.1, F33.2, F33.4, F33.8, F33.9) documented in the medical records at the time of enrollment, age 18-45 y, having no established mindfulness practice, having a smartphone with iOS/ Android system, and willingness to commit to the intervention. Exclusion criteria included ongoing psychotherapy, active substance abuse, severe suicidal ideation, previous suicide attempt, and severe comorbid mental disorder such as psychosis.

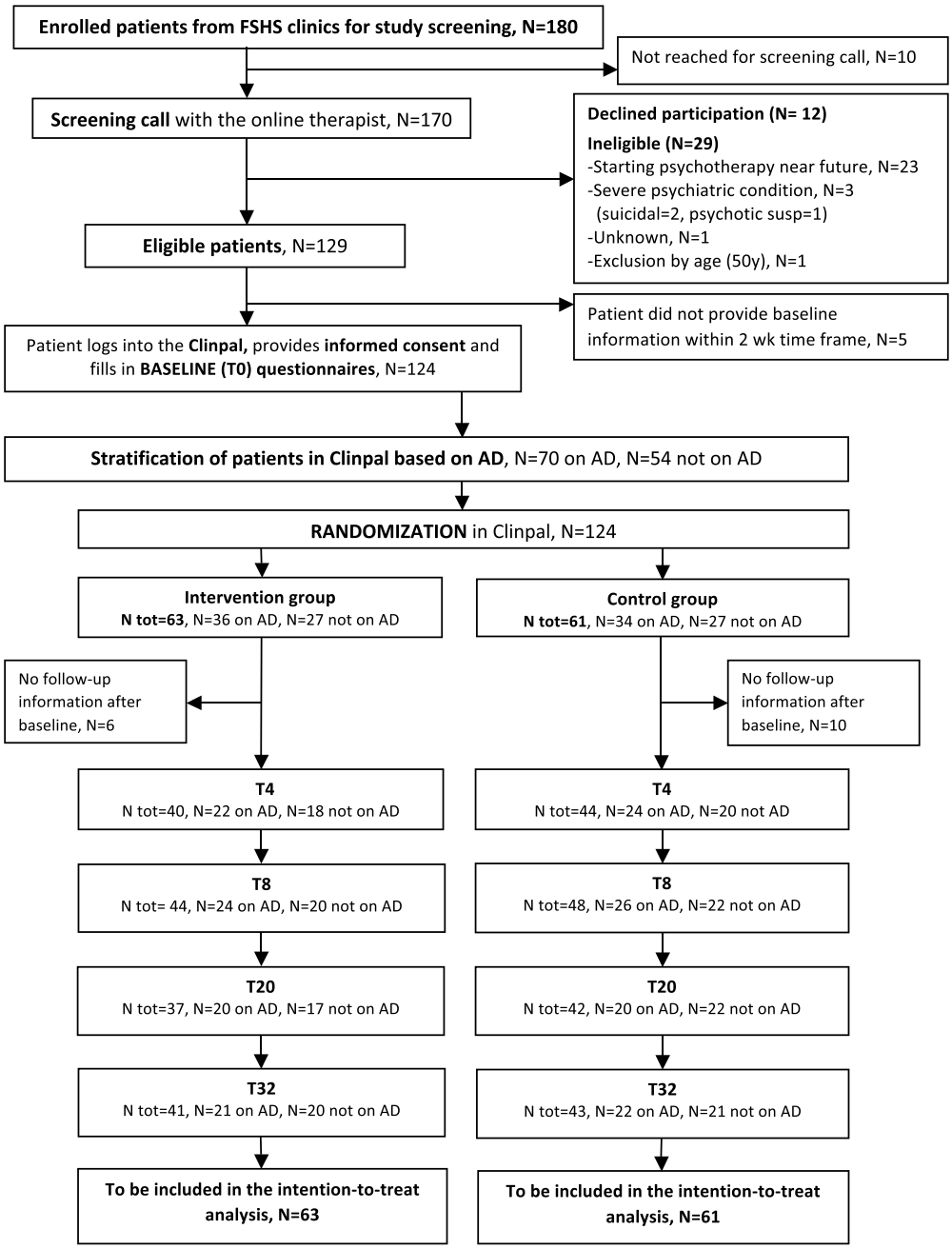

Fig. 1. CONSORT flow diagram for a randomized clinical trial of therapist-guided smartphone app for major depression in young adults.
FSHS, Finnish Student Health Service; **AD**, Antidepressant medication; **T0**, baseline, before the intervention start; **T4**, 4 weeks from the baseline (mid-intervention); **T8**, 8 weeks from the baseline (intervention completion); **T20**, 20 weeks from the baseline (3 month follow-up after the intervention completion); **T32**, 32 weeks from the baseline (6 month follow-up after the intervention completion).

2.4. Measures

The following self-reported instruments were addressed at baseline (T0) and every follow-up (T4, T8, T20, T32), except for the System Usability Scale (SUS), which was administered only at the post-intervention follow-up (T8).

2.4.1. Patient Health Questionnaire 9-point (PHQ-9)

The primary outcome measure was Patient Health Questionnaire-9 (PHQ-9) for depression symptoms (Löwe et al., 2004); PHQ-9 has demonstrated excellent internal consistency reliability, with Cronbach's alpha of 0.89 in primary care settings, and excellent test-retest reliability (Kroenke et al., 2001).

2.4.2. Generalized Anxiety Disorder 7-item (GAD-7)

Generalized Anxiety Disorder-7 (GAD-7), a secondary outcome measure, was used to measure for anxiety symptoms (Löwe et al., 2008). The instrument has shown an excellent internal consistency in the general population with Cronbach's alpha of 0.89 (Löwe et al., 2008).

2.4.3. Insomnia Severity Index (ISI)

Insomnia Severity Index (ISI) (Morin et al., 2011), a secondary outcome measure, was used to measure sleep disturbance. ISI has been shown to be reliable and valid tool of measuring insomnia in the general population and assessing treatment response in patient populations. It has demonstrated excellent internal consistency, with Cronbach's alpha of 0.90-0.91 (Morin et al., 2011).

2.4.4. EUROHIS-Qol 8-item index (EUROHIS-Qol-8)

EUROHIS-Qol 8-item index (EUROHIS-Qol-8), a secondary outcome measure, was used to measure quality of life (Schmidt et al., 2006). EUROHIS-Qol 8-item index has demonstrated a good internal

Table 1
Participant characteristics at baseline (T0).

	All	Intervention group	Control group	Difference between intervention and control groups, p
N	124	63	61	
Age (y), Mean (SD, range)	25.1 (4.5)	24.5 (3.4, 19-36 y)	25.8 (5.4, 19-44 y)	0.11
Sex				0.19
Female, *n* (%)	90 (72.6%)	49 (77.8%)	41 (67.2%)	
Male, *n* (%)	34 (27.4%)	14 (22.2%)	20 (32.8%)	
Antidepressant medication				0.88
Yes, *n* (%)	70 (56.5%)	36 (57.1%)	34 (55.7%)	
No, *n* (%)	54 (43.5%)	27 (42.9%)	27 (44.3%)	
Social support				0.58
Weak *n* (%)	9 (8.0%)	6 (10.2%)	3 (5.6%)	
Moderate *n* (%)	25 (22.1%)	14 (23.7%)	11 (20.4%)	
Strong *n* (%)	79 (69.9%)	39 (66.1%)	40 (74.1%)	
Missing information, *n* (%)	7 (5.6%)	4 (6.3%)	3 (4.9%)	
Baseline scores, Mean (SD)				
PHQ-9	12.0 (4.4)	12.4 (4.2)	11.6 (4.6)	0.26
GAD-7	9.5 (4.1)	9.6 (3.9)	9.4 (4.3)	0.84
ISI	9.9 (4.9)	9.7 (4.7)	10.1 (5.1)	0.65
EUROHIS-Qol-8	23.4 (3.8)	22.9 (3.6)	23.8 (4.0)	0.17
FFMQ-SF	68.1 (9.3)	67.0 (8.9)	69.2 (9.5)	0.18
PSS-10	25.3 (5.2)	26.1 (4.7)	24.5 (5.7)	0.09
Resilience Scale	76.8 (11.4)	75.5 (12.3)	77.8 (10.3)	0.27

Abbreviations: PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; ISI, Insomnia Severity Index; EUROHIS-Qol-8, EUROHIS-Qol 8-item index; FFMQ-SF, Five Facet Mindfulness Questionnaire-Short Form; PSS-10, Perceived Stress Scale.

consistency with Cronbach's alpha of 0.83 (Schmidt et al., 2006).

2.4.5. Five Facet Mindfulness Questionnaire–Short Form (FFMQ-SF)

Five Facet Mindfulness Questionnaire - Short Form (FFMQ-SF) (Bohlmeijer et al., 2011), a secondary outcome measure, was used to measure the internalization of mindfulness skills. Internal consistency of FFMQ-SF has been shown to be good with Cronbach's alpha ranging from 0.77 to 0.93 (Williams et al., 2014).

2.4.6. Perceived Stress Scale (PSS-10)

Perceived Stress Scale (PSS-10) (Cohen et al., 1983), a secondary outcome measure, was used to measure the level of perceived psychological stress. PSS-10 has demonstrated good internal consistency with Cronbach's alpha ranging from 0.78 to 0.91 (Lee, 2012).

2.4.7. Resilience Scale

Resilience Scale (Wagnild and Young, 1993), a secondary outcome measure, was used to measure resilience. The Resilience Scale has demonstrated good internal consistency with Cronbach's alpha ranging from 0.76 to 0.91 (Wagnild and Young, 1993).

2.4.8. System Usability Scale (SUS)

System Usability Scale (SUS) (Jordan et al., 1996), a secondary outcome measure, was used to measure the usability of the mobile app among patients in the intervention group. SUS has demonstrated excellent internal consistency with Cronbach's alpha of 0.92 (Lewis, 2018).

2.4.9. The intervention engagement

Total engagement time (frequency x duration) to the intervention mindfulness practices and chat activity (total message count), i.e. participants' engagement with a therapist and vice versa, were measured automatically in the smartphone app.

2.4.10. Social support

Social support was assessed by a study therapist by an informal scale, weak/ moderate/strong, based on the baseline information.

2.5. Intervention

MHP is a comprehensive therapist-guided intervention for depression delivered via a smartphone app (Economides et al., 2019). In the current trial, we used 8-week version of the MHP, which consists of 8 sequentially delivered modules, with content derived from Mindfulness-Based Stress Reduction (Kabat-Zinn, 1982), Mindfulness-Based Cognitive Therapy (Segal et al., 2002), CBT (Beck, 1976), and Behavioral Activation Therapy (Jacobson et al., 2001). The content includes text, videos, audio-guided mindfulness exercises, infographics illustrating CBT principles, and journal prompts. Daily content and practice time range between 10-45 min. MHP includes anonymous peer support via moderated group discussion board, and asynchronous support by a remote therapist. Therapists are board-certified and licensed master's level mental health professionals, who review participant engagement and provide one-to-one support via chat messaging, and infrequently, by phone calls. If there were signs of patient's mental state deterioration, therapist conducted additional phone-based assessment.

2.6. Treatment as usual (TAU)

TAU comprised of pragmatic treatment for depression as provided at the FSHS. TAU could or could not include AD, laboratory tests, and appointments with healthcare professionals such as nurses, psychologists, and doctors. Frequency of the appointments was tailored to each patient's individual needs.

2.7. Statistical analysis

Assuming 85% power, α-level of 0.05, and allowing 35% drop-out, a mean difference of 5 points (SD 5.2) in primary outcome measure between the intervention and the control group could be expected to be detected with 30 patients in each stratified group, which yielded 60 patients per arm (*n* total 120). Baseline differences in outcomes between intervention and control groups were tested with two-sample t-test. Differences in the changes in primary and secondary outcomes between intervention and control groups and the changes within groups were analyzed with a restricted maximum likelihood estimation method based repeated measures model using the PROC MIXED in SAS. The model included the effects of group, week and group*week interaction. Least squares means (LSM) for changes (from baseline to follow-up) within groups and for the difference in the changes between groups were estimated using contrasts in the repeated measures model. Follow-up measurements were compared to the baseline using Dunnett's adjustment in pairwise comparisons. Analyses followed intent to treat approach under a missing at random assumption. As a sensitivity analysis, last observation carried forward (LOCF) method was used to impute the missing values. The effect sizes between the intervention and the

Table 2
Comparison between intervention and control groups. Results of repeated measures analysis of variance; least square means and least square mean changes at 4, 8, 20 and 32 weeks compared to the baseline with Dunnett's adjustment method. Effect sizes (*Cohen's d*) were calculated as the difference in mean change between intervention and control groups divided by pooled standard deviation at the baseline.

	Intervention			Control					
	Mean values (SE)	Mean change (95% CI)	*P**	Mean values (SE)	Mean change (95% CI)	*P**	Mean difference (95% CI)	Group *Time interaction p	Intervention effect size, Cohen's *d*
PHQ-9									
T0, baseline	12.44 (0.58)			11.56 (0.59)					
T4, 4 weeks	10.58 (0.67)	-1.86 (-3.47 to -0.25)	0.017	8.88 (0.66)	-2.68 (-4.24 to -1.12)	0.0001	0.82 (-0.97 to 2.61)	0.37	0.19
T8, 8 weeks	9.89 (0.65)	-2.55 (-4.11 to -1.00)	0.0002	8.57 (0.64)	-2.99 (-4.51 to -1.48)	<0.0001	0.44 (-1.29 to 2.17)	0.62	0.10
T20, 20 weeks	8.88 (0.69)	-3.57 (-5.22 to -1.91)	<0.0001	8.68 (0.67)	-2.88 (-4.47 to -1.29)	<0.0001	-0.69 (-2.51 to 1.14)	0.46	-0.16
T32, 32 weeks	8.09 (0.67)	-4.36 (-5.95 to -2.76)	<0.0001	8.62 (0.66)	-2.94 (-4.51 to -1.36)	<0.0001	-1.42 (-3.21 to 0.36)	0.12	-0.32
GAD-7									
T0, baseline	9.59 (0.52)			9.44 (0.52)					
T4, 4 weeks	8.95 (0.61)	-0.64 (-2.15 to 0.87)	0.70	7.23 (0.59)	-2.21 (-3.67 to -0.75)	0.0009	1.57 (-0.11 to 3.24)	0.07	0.38
T8, 8 weeks	7.86 (0.59)	-1.72 (-3.18 to -0.26)	0.014	6.91 (0.57)	-2.53 (-3.96 to -1.11)	<0.0001	0.81 (-0.81 to 2.43)	0.33	0.20
T20, 20 weeks	6.81 (0.62)	-2.78 (-4.33 to -1.23)	<0.0001	7.05 (0.60)	-2.39 (-3.88 to -0.91)	0.0003	-0.39 (-2.10 to 1.33)	0.66	-0.10
T32, 32 weeks	6.52 (0.60)	-3.07 (-4.57 to -1.57)	<0.0001	7.11 (0.60)	-2.33 (-3.81 to -0.86)	0.0005	-0.74 (-2.41 to 0.94)	0.39	-0.18
ISI									
T0, baseline	9.68 (0.61)			10.08 (0.62)					
T4, 4 weeks	8.68 (0.69)	-1.00 (-2.51 to 0.50)	0.31	9.21 (0.68)	-0.87 (-2.33 to 0.59)	0.41	-0.13 (-1.80 to 1.54)	0.88	-0.03
T8, 8 weeks	8.42 (0.68)	-1.26 (-2.71 to 0.20)	0.11	7.85 (0.67)	-2.23 (-3.65 to -0.81)	0.0005	0.97 (-0.65 to 2.59)	0.24	0.20
T20, 20 weeks	8.26 (0.71)	-1.43 (-2.98 to 0.12)	0.08	8.12 (0.69)	-1.96 (-3.44 to -0.47)	0.005	0.53 (-1.18 to 2.24)	0.54	0.11
T32, 32 weeks	7.86 (0.69)	-1.82 (-3.31 to -0.33)	0.01	7.49 (0.69)	-2.59 (-4.07 to -1.12)	<0.0001	0.77 (-0.90 to 2.44)	0.37	0.16
EUROHIS-Qol-8									
T0, baseline	22.89 (0.56)			23.82 (0.57)					
T4, 4 weeks	24.18 (0.64)	1.29 (-0.14 to 2.72)	0.09	25.49 (0.63)	1.67 (0.28 to 3.06)	0.012	-0.38 (-1.97 to 1.21)	0.64	-0.10
T8, 8 weeks	24.82 (0.62)	1.93 (0.55 to 3.32)	0.003	26.19 (0.61)	2.37 (1.02 to 3.72)	<0.0001	-0.44 (-1.98 to 1.10)	0.58	-0.12
T20, 20 weeks	24.96 (0.65)	2.07 (0.59 to 3.54)	0.002	26.67 (0.63)	2.85 (1.44 to 4.27)	<0.0001	-0.79 (-2.42 to 0.84)	0.34	-0.21
T32, 32 weeks	26.13 (0.63)	3.24 (1.82 to 4.66)	<.0001	26.24 (0.63)	2.42 (1.02 to 3.83)	0.0001	0.81 (-0.78 to 2.40)	0.32	0.21
FFMQ-SF									
T0, baseline	66.97 (1.26)			69.21 (1.28)					
T4, 4 weeks	68.62 (1.42)	1.66 (-1.38 to 4.69)	0.49	71.32 (1.40)	2.11 (-0.84 to 5.05)	0.24	-0.45 (-3.82 to 2.92)	0.79	-0.05
T8, 8 weeks	73.19 (1.39)	6.22 (3.29 to 9.15)	<0.0001	72.59 (1.37)	3.38 (0.52 to 6.24)	0.014	2.84 (-0.42 to 6.11)	0.09	0.31
T20, 20 weeks	74.61 (1.45)	7.64 (4.52 to 10.77)	<0.0001	73.12 (1.42)	3.91 (0.91 to 6.90)	0.005	3.74 (0.29 to 7.18)	0.034	0.40
T32, 32 weeks	76.21 (1.41)	9.24 (6.23 to 12.25)	<0.0001	73.15 (1.41)	3.94 (0.97 to 6.91)	0.005	5.30 (1.93 to 8.67)	0.002	0.57
PSS-10									
T0, baseline	26.06 (0.71)			24.48 (0.72)					
T4, 4 weeks	23.87 (0.81)	-2.19 (-4.00 to -0.38)	0.011	22.87 (0.79)	-1.60 (-3.36 to 0.16)	0.09	-0.59 (-2.60 to 1.42)	0.57	-0.11
T8, 8 weeks	22.35 (0.78)	-3.71 (-5.47 to -1.96)	<0.0001	21.96 (0.77)	-2.52 (-4.22 to -0.81)	0.001	-1.20 (-3.15 to 0.75)	0.23	-0.23
T20, 20 weeks	21.27 (0.83)	-4.88 (-6.65 to -2.92)	<0.0001	21.77 (0.80)	-2.70 (-4.49 to -0.92)	0.0009	-2.08 (-4.14 to -0.03)	0.047	-0.40
T32, 32 weeks	20.83 (0.80)	-5.23 (-7.03 to -3.44)	<0.0001	21.96 (0.80)	-2.52 (-4.29 to -0.74)	0.002	-2.71 (-4.73 to -0.70)	0.008	-0.52
Resilience Scale									

(*continued on next page*)

Table 2 (*continued*)

	Intervention			Control					
	Mean values (SE)	Mean change (95% CI)	P*	Mean values (SE)	Mean change (95% CI)	P*	Mean difference (95% CI)	Group *Time interaction p	Intervention effect size, Cohen's *d*
T0, baseline	75.51 (1.48)			77.75 (1.50)					
T4, 4 weeks	78.95 (1.60)	3.44 (0.57 to 6.30)	0.012	80.72 (1.59)	2.97 (0.18 to 5.75)	0.033	0.47 (-2.71 to 3.66)	0.77	0.04
T8, 8 weeks	81.82 (1.57)	6.31 (3.54 to 9.09)	<0.0001	82.32 (1.57)	4.56 (1.86 to 7.27)	0.0002	1.75 (-1.33 to 4.84)	0.27	0.15
T20, 20 weeks	83.22 (1.63)	7.71 (4.76 to 10.66)	<0.0001	82.89 (1.61)	5.14 (2.30 to 7.97)	<0.0001	2.58 (-0.68 to 5.83)	0.12	0.23
T32, 32 weeks	84.71 (1.59)	9.20 (6.36 to 12.05)	<0.0001	83.33 (1.60)	5.57 (2.76 to 8.38)	<0.0001	3.63 (0.44 to 6.82)	0.026	0.32

Abbreviations: SE, Standard Error; CI, Confidence Interval; PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; ISI, Insomnia Severity Index; EUROHIS-Qol-8, EUROHIS-Qol 8-item index; FFMQ-SF, Five Facet Mindfulness Questionnaire - Short Form; PSS-10, Perceived Stress Scale.

control group were calculated by *Cohen's d*. Effect sizes of 0.2 were considered as small, 0.5 as medium, and 0.8 as large. Clinical significance for the primary outcome was defined as ≥5 points' reduction in PHQ-9 score (Löwe et al., 2004). Binary logistic regression was applied to compare ≥5 points' decrease in PHQ-9 score at T8 and T32 between the groups. Relationship between the amount of mindfulness practice and the change of depression and anxiety symptoms among the intervention group was examined using Pearson correlation coefficients. The potential modifying effect of internalization of mindfulness skills by FFMQ-SF (divided into tertiles) on the association of the quantity of mindfulness practice and the PHQ-9 change and GAD-7 change were examined using linear model. The potential modifying effect of AD on intervention effect was tested post-hoc, and exploratory subgroup analyses, including primary and secondary 1 outcome analyses, by AD status, were done. Statistical analyses were performed with SAS software (SAS software). Two-sided statistical tests with 0.05 level of significance were used.

3. Results

3.1. Sample and participant characteristics

Altogether 180 patients were enrolled, and 170 were screened; of these, 12 declined to participate, 29 were ineligible, 5 did not provide baseline information, and 124 were randomized – 63 to receive MHP plus TAU and 61 to receive TAU. Drop-out was defined as having provided no follow-up information after the baseline (attrition 12.8%, n = 10 control, n = 6 intervention, p = 0.27). The CONSORT diagram is presented in Fig. 1. Participants' baseline characteristics are summarized in Table 1; the differences between the two groups suggest no significant effect of randomization.

3.2. Primary outcome

There was no statistically or clinically significant difference between the intervention group and the control group in the reduction in depression symptoms by PHQ-9 from baseline to immediately post-intervention, T8 (LSM difference 0.44, *95% CI* = -1.29, 2.17, p = 0.62, *Cohen's d* = 0.10), or to T32 (LSM difference -1.42, *95% CI* = -3.21, 0.36, p =0.12, *Cohen's d* = -0.32) (Table 2). At T8, 27.0% (17/63) in the intervention group versus 26.2% (16/61) in the control group achieved ≥5 points' decrease in PHQ-9 (LOCF-analysis, *OR* = 1.04, *95% CI* = 0.47, 2.31, p = 0.92). At T32, the respective was true for 39.7% (25/63) in the intervention group versus 26.2% (16/61) in the control group (LOCF-analysis, *OR* = 1.85, *95% CI* = 0.86, 3.96, p = 0.11).

3.3. Secondary outcomes

No statistically significant differences between the intervention group and the control group in change of anxiety by GAD-7, sleep disturbances by ISI, and quality of life by EUROHIS-Qol-8 were detected (Table 2). The increase in mindfulness by FFMQ-SF was larger in the intervention group than in the control group at T20 and at T32 (LSM difference 5.30, *95% CI* = 1.93, 8.67, p = 0.002, *Cohen's d* = 0.57). The reduction of the perceived stress by PSS-10 was larger in the intervention group than in the control group at T20 and at T32 (LSM difference -2.71, *95% CI* = -4.73, -0.70, p = 0.008, *Cohen's d* = -0.52). The increase in Resilience Scale was larger in the intervention group than in the control group at T32 (LSM difference 3.63, *95% CI* = 0.44, 6.82, p = 0.03, *Cohen's d* = 0.32) (Table 2).

The mean total minutes of the intervention-associated mindfulness practice was 496 (*SD* = 274, median 501, range 29 - 1161), equaling 8.9 minutes of daily practice during the intervention. The associations between total minutes of completed mindfulness practices and reduction of depression by PHQ-9 at T8 (r = -0.22, p = 0.15), T20 (r = -0.08, p = 0.63) and T32 (r = -0.0002, p = 0.99) were not significant. The same was true for anxiety by GAD-7 at T8 (r = -0.16, p = 0.29), T20 (r = -0.02, p = 0.89) and T32 (r = 0.11, p = 0.51).

The level of mindfulness skills by FFMQ-SF did not significantly modify the association of the total time of completed mindfulness practice and change of depression symptoms by PHQ-9 at T8 (p = 0.91), T20 (p = 0.18), or T32 (p = 0.19) or change of GAD-7 at T8 (p = 0.41) or T20 (p = 0.06). At T32, FFMQ-SF modified the association of the amount of mindfulness practice and change of GAD-7 score (p = 0.017): higher total minutes were associated with the reduction of anxiety by GAD-7 in the lowest tertile (β = -0.014, *SE* = 0.006, p = 0.048). In the medium tertile, however, higher amount of practice was marginally significantly associated with the increase of anxiety by GAD-7 (β = 0.012, SE = 0.006, p = 0.054), and no association was detected in the highest tertile (β = 0.002, *SE* = 0.004, p = 0.58).

3.4. Sensitivity analysis

In the sensitivity analyses (LOCF) of primary and secondary outcome 1 analyses (Table S2), mindfulness by FFMQ-SF was no longer significant at T20. Otherwise results of the sensitivity analyses were consistent with the results of the main analyses.

3.5. Intervention effects by antidepressant medication (AD)

To assess whether the intervention effects were modified by AD, exploratory analysis of the interaction AD x intervention x time were conducted. There was a trend for the interaction on primary outcome, depression by PHQ-9 (AD x group x time effect, p = 0.10). Primary and secondary outcome 1 analyses were therefore conducted separately for those on AD and those not on AD.

No significant differences between the intervention and the control group were seen in the baseline scores among patients on AD (n = 70)

Table 3
Comparison between intervention and control groups on antidepressant medication. Results of repeated measures analysis of variance; least square means and least square mean changes at 4, 8, 20 and 32 weeks compared to the baseline with Dunnett's adjustment method. Effect sizes (*Cohen's d*) were calculated as the difference in mean change between intervention and control groups divided by pooled standard deviation at the baseline.

	Intervention group on AD (*n*= 36)			Control group on AD (*n*= 34)					
	Mean values (SE)	Mean change (95% CI)	*P*	Mean values (SE)	Mean change (95% CI)	*P**	Mean difference (95% CI)	Group *Time interaction, p	Intervention effect size, Cohen's *d*
PHQ-9									
T0, baseline	13.56 (0.81)			12.65 (0.83)					
T4, 4 weeks	12.39 (0.95)	-1.17 (-3.43 to 1.10)	0.54	10.58 (0.93)	-2.07 (-4.28 to 0.14)	0.07	0.90 (-1.61 to 3.41)	0.48	0.20
T8, 8 weeks	11.17 (0.92)	-2.39 (-4.58 to -0.20)	0.028	9.61 (0.91)	-3.04 (-5.19 to -0.89)	0.002	0.65 (-1.79 to 3.09)	0.60	0.14
T20, 20 weeks	9.36 (0.98)	-4.19 (-6.54 to -1.85)	<0.0001	10.20 (0.99)	-2.45 (-4.80 to -0.10)	0.038	-1.74 (-4.38 to 0.90)	0.20	-0.38
T32, 32 weeks	8.02 (0.97)	-5.54 (-7.84 to -3.24)	<0.0001	10.46 (0.96)	-2.19 (-4.47 to 0.09)	0.06	-3.35 (-5.92 to -0.77)	0.011	-0.73
GAD-7									
T0, baseline	9.56 (0.71)			10.23 (0.73)					
T4, 4 weeks	9.39 (0.84)	-0.16 (-2.18 to 1.86)	0.999	7.58 (0.83)	-2.66 (-4.63 to -0.68)	0.004	2.50 (0.25 to 4.74)	0.030	0.59
T8, 8 weeks	7.85 (0.82)	-1.70 (-3.66 to 0.26)	0.11	7.29 (0.81)	-2.94 (-4.87 to -1.02)	0.0008	1.24 (-0.94 to 3.43)	0.26	0.29
T20, 20 weeks	6.55 (0.87)	-3.00 (-5.10 to -0.91)	0.002	7.21 (0.88)	-3.03 (-5.13 to -0.92)	0.002	0.02 (-2.34 to 2.39)	0.98	0.00
T32, 32 weeks	6.26 (0.85)	-3.29 (-5.35 to -1.23)	0.0004	8.03 (0.85)	-2.21 (-4.24 to -0.17)	0.029	-1.09 (-3.39 to 1.22)	0.35	-0.26
ISI									
T0, baseline	11.14 (0.83)			10.41 (0.85)					
T4, 4 weeks	10.47 (0.96)	-0.67 (-2.83 to 1.49)	0.88	10.67 (0.95)	0.26 (-1.85 to 2.38)	0.99	-0.93 (-3.34 to 1.47)	0.45	-0.20
T8, 8 weeks	9.54 (0.94)	-1.60 (-3.70 to 0.50)	0.20	8.85 (0.93)	-1.56 (-3.62 to 0.49)	0.20	-0.04 (-2.37 to 2.30)	0.98	-0.01
T20, 20 weeks	9.76 (0.99)	-1.38 (-3.62 to 0.87)	0.38	9.21 (1.00)	-1.20 (-3.45 to 1.05)	0.50	-0.17 (-2.70 to 2.35)	0.89	-0.04
T32, 32 weeks	8.10 (0.97)	-3.03 (-5.24 to -0.83)	0.003	8.41 (0.97)	-2.00 (-4.18 to 0.18)	0.08	-1.04 (-3.50 to 1.43)	0.41	-0.22
EUROHIS-Qol-8									
T0, baseline	22.31 (0.78)			23.38 (0.80)					
T4, 4 weeks	22.67 (0.89)	0.37 (-1.61 to 2.34)	0.98	24.46 (0.88)	1.08 (-0.85 to 3.01)	0.46	-0.71 (-2.91 to 1.49)	0.52	-0.17
T8, 8 weeks	23.76 (0.87)	1.45 (-0.46 to 3.37)	0.20	25.82 (0.87)	2.44 (0.56 to 4.32)	0.006	-0.98 (-3.12 to 1.15)	0.36	-0.24
T20, 20 weeks	23.71 (0.92)	1.41 (-0.64 to 3.46)	0.28	25.50 (0.93)	2.12 (0.06 to 4.17)	0.042	-0.71 (-3.02 to 1.60)	0.55	-0.17
T32, 32 weeks	24.80 (0.91)	2.49 (0.48 to 4.51)	0.009	24.62 (0.91)	1.24 (-0.75 to 3.23)	0.36	1.25 (-1.00 to 3.51)	0.27	0.30
FFMQ-SF									
T0, baseline	66.75 (1.80)			69.26 (1.85)					
T4, 4 weeks	68.66 (2.05)	1.91 (-2.49 to 6.31)	0.69	70.67 (2.03)	1.41 (-2.89 to 5.70)	0.85	0.50 (-4.39 to 5.39)	0.84	0.05
T8, 8 weeks	73.18 (2.00)	6.43 (2.16 to 10.69)	0.001	72.18 (2.0)	2.92 (-1.26 to 7.10)	0.26	3.51 (-1.24 to 8.26)	0.15	0.35
T20, 20 weeks	73.66 (2.11)	6.91 (2.35 to 11.47)	0.0009	71.86 (2.13)	2.60 (-1.98 to 7.18)	0.45	4.31 (-0.83 to 9.45)	0.10	0.43
T32, 32 weeks	76.95 (2.08)	10.20 (5.72 to 14.67)	<0.0001	72.03 (2.08)	2.76 (-1.67 to 7.19)	0.36	7.43 (2.42 to 12.44)	0.004	0.74
PSS-10									
T0, baseline	25.97 (1.00)			25.09 (1.03)					
T4, 4 weeks	24.57 (1.13)	-1.40 (-3.70 to 0.90)	0.39	24.06 (1.12)	-1.02 (-3.27 to 1.23)	0.64	-0.38 (-2.94 to 2.18)	0.77	-0.07
T8, 8 weeks	22.94 (1.11)	-3.03 (-5.26 to -0.80)	0.004	22.21 (1.11)	-2.88 (-5.07 to -0.69)	0.005	-0.15 (-2.64 to 2.34)	0.90	-0.03
T20, 20 weeks	21.75 (1.16)	-4.22 (-6.61 to -1.83)	<0.0001	22.81 (1.17)	-2.28 (-4.68 to 0.11)	0.07	-1.94 (-4.63 to 0.76)	0.16	-0.35
T32, 32 weeks	20.30 (1.14)	-5.68 (-8.02 to -3.33)	<0.0001	23.44 (1.15)	-1.65 (-3.97 to 0.67)	0.24	-4.02 (-6.65 to -1.40)	0.003	-0.72
Resilience Scale									

(*continued on next page*)

Table 3 (*continued*)

	Intervention group on AD (*n*= 36)			Control group on AD (*n*= 34)					
	Mean values (SE)	Mean change (95% CI)	*P*	Mean values (SE)	Mean change (95% CI)	*P**	Mean difference (95% CI)	Group *Time interaction, p	Intervention effect size, Cohen's *d*
T0, baseline	72.86 (2.09)			77.44 (2.15)					
T4, 4 weeks	76.41 (2.26)	3.55 (-0.22 ro 7.32)	0.07	79.53 (2.28)	2.09 (-1.59 to 5.77)	0.44	1.46 (-2.73 to 5.65)	0.49	0.12
T8, 8 weeks	78.25 (2.23)	5.39 (1.73 to 9.04)	0.001	81.40 (2.25)	3.96 (0.37 to 7.54)	0.025	1.43 (-2.65 to 5.50)	0.49	0.12
T20, 20 weeks	80.38 (2.30)	7.52 (3.61 to 11.43)	<0.0001	82.22 (2.34)	4.78 (0.86 to 8.71)	0.011	2.73 (-1.68 to 7.14)	0.22	0.22
T32, 32 weeks	83.47 (2.28)	10.60 (6.76 to 14.44)	<0.0001	82.03 (2.31)	4.58 (0.78 to 8.38)	0.012	6.02 (1.72 to 10.32)	0.006	0.49

Abbreviations: SE, Standard Error; CI, Confidence Interval; AD, Antidepressant medication; PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; ISI, Insomnia Severity Index; EUROHIS-Qol-8, EUROHIS-Qol 8-item index; FFMQ-SF, Five Facet Mindfulness Questionnaire-Short Form; PSS-10, Perceived Stress Scale.

and not on AD at baseline ($n = 54$) (Table S1). The intervention group on AD had a statistically significantly greater reduction in depression by PHQ-9 from baseline to T32 (LSM difference -3.35, *95% CI* = -5.92, -0.77, $p = 0.01$, *Cohen's d* = -0.73) compared to the control group on AD (Table 3). Decrease of ≥5 points' in PHQ-9 score in those on AD was achieved by 22.2% (8/36) in the intervention group versus 26.5% (9/34) in the control group at T8 (LOCF-analysis, *OR* = 0.79, *95% CI* = 0.27, 2.37, $p = 0.68$), and by 38.9% (14/36) in the intervention group versus 26.5% (9/34) in the control group at T32 (LOCF-analysis, *OR* = 1.77, *95% CI* = 0.64, 4.88, p = 0.27). Control group on AD had statistically significantly greater reduction in anxiety by GAD-7 compared to the intervention group on AD from baseline to T4 (LSM difference 2.50, *95% CI* = 0.25, 4.74, $p = 0.03$, *Cohen's d* = 0.59). Intervention group on AD had greater reduction in stress by PSS-10 compared to controls on AD from baseline to T32 (LSM difference -4.02, *95% CI* = -6.65, -1.40, $p = 0.003$, *Cohen's d* = -0.72). Respectively, from baseline to T32, there was an increase among intervention group on AD in mindfulness by FFMQ-SF (LSM difference 7.43, *95% CI* = 2.42, 12.44, $p = 0.004$, *Cohen's d* = 0.74), and in resilience by Resilience Scale (LSM difference 6.02, *95% CI* = 1 .72, 10.32, $p = 0.006$, *Cohen's d* = 0.49) (Table 3). Among those not on AD (Table 4), comparing the intervention group to the control group, controls had greater decrease of sleep disturbances by ISI at T8 and at T32 (LSM difference 2.82, *95% CI* = 0.61, 5.03, $p = 0.013$, *Cohen's d* = 0.59). Decrease of ≥5 points' in PHQ-9 in those not on AD was achieved by 33.3% (9/27) in the intervention group versus 25.% (7/27) in the control group at T8 (LOCF-analysis, *OR* = 1.43, *95% CI* = 0.44, 4.63, $p = 0.55$), and by 40.7% (11/27) in the intervention group versus 25.9% (7/27) in the control group at T32 (LOCF-analysis, *OR* = 1.40, *95% CI* 0.45, 4.36, $p = 0.56$). Results of the sensitivity analyses at T4, T8, T20 and T32 in those on and not on AD were consistent with the results from the main analyses (data not shown).

3.6. App use and adverse events

The mean SUS score of the MHP was 86.9 (median 87.5, range 60.0-100.0), suggesting above average usability (cut-off 68 points) among respondents ($n = 44$, 69.8% of the intervention group). Chat activity, i.e. messages during the intervention from patient to therapist (mean 22, median 19, range 1-77), or from therapist to patient (mean 54, median 55, range 17-122) were not associated with the change of depression by PHQ-9. There were no study-related adverse events.

4. Discussion

The overall aim of this randomized trial was to evaluate the efficacy of an app-based mobile intervention, MHP, plus TAU compared to TAU in patients with clinical depression recruited from the Finnish university students' health services. The results showed that the intervention group receiving 8-week MHP plus TAU and the control group receiving only TAU did not differ significantly from one another on depression symptoms, either from baseline to immediately post-intervention or to the 6-month follow-up after the intervention. The therapist-guided mobile app did thus not provide additional benefit over TAU alone for the treatment of depression among young adults representing highly educated population. Within the intervention group, reduction of PHQ-9 score was 4.4, and the difference to controls was 1.4 (*d* = -0.32), clearly below the minimal clinically important difference for the measure. Of secondary outcomes, we observed significant improvement among intervention group in mindfulness skills, which is unsurprising given that mindfulness is a central method in MHP; but additionally, the intervention group demonstrated significant improvement in resilience and reduction in experienced stress, i.e. in qualities highly relevant for recovering from depression and for protection of depressive relapse (Strain, 2018). The time spent on intervention-associated mindfulness practices did not associate with the change of depression or anxiety, suggesting that at least in the examined range, increasing mindfulness practice time was not associated with improved outcomes. Results also suggested that internalization of the mindfulness skills did not modify the effect of the mindfulness practice on the change of depression symptoms, which implies that internalization of the mindfulness does not influence the relationship between the amount of practice and the change of depression. In anxiety however, total time of the mindfulness practice was associated with the reduction of anxiety among those with lower mindfulness skills at the end of the follow-up. This suggests that in the long run, increasing the mindfulness practice time has relevance for the improvement of anxiety among those with lower but not higher mindfulness skills.

AD seemed to modify the intervention effect on primary outcome, depression, and overall, patients on AD benefitted from the MHP more than those not on AD. Intervention effect size for depression symptoms in those on AD was moderate, whereas in the whole intervention group it was small and non-significant, and in those not on AD, it was non-existent. Within the intervention group on AD, reduction of PHQ-9 score was 5.5; the difference to controls was 3.4, which is below clinical significance (≥ 5 points) as defined in the current trial. However, Löwe et al. (2004) have also suggested range 3 - 5 points for the minimal clinically important difference of PHQ-9, which would allow interpretation that the difference to control group was clinically significant among those on AD. Significant differences between the patients on AD in the intervention and the control group, at the direction of hypotheses, were also apparent in mindfulness skills, experienced stress, and resilience. The hypothesized pattern that the maximum intervention benefit would manifest immediately post-intervention (T8), and decrease slightly during the 6-month follow-up, turned out not be true among those on AD; instead, the effects of the intervention on all outcomes increased as a function of time. In contrast to patients on AD, the only difference between the intervention and the control group among

Table 4
Comparison between intervention and control groups not on antidepressant medication. Results of repeated measures analysis of variance; least square means and least square mean changes at 4, 8, 20 and 32 weeks compared to the baseline with Dunnett's adjustment method. Effect sizes (*Cohen's d*) were calculated as the difference in mean change between intervention and control groups divided by pooled standard deviation at the baseline.

	Intervention group not on AD (n= 27)			Control group not on AD (n= 27)					
	Mean values (SE)	Mean change (95% CI)	P*	Mean values (SE)	Mean change (95% CI)	P*	Mean difference (SE)	Group *Time interaction, p	Intervention effect size, Cohen's *d*
PHQ-9									
T0, baseline	10.96 (0.73)			10.19 (0.73)					
T4, 4 weeks	8.31 (0.86)	-2.65 (-4.93 to -0.37)	0.016	6.84 (0.82)	-3.34 (-5.54 to -1.15)	0.0009	0.69 (-1.83 to 3.21)	0.59	0.18
T8, 8 weeks	8.27 (0.82)	-2.69 (-4.89 to -0.49)	0.011	7.31 (0.80)	-2.87 (-5.00 to -0.74)	0.004	0.18 (-2.25 to 2.62)	0.88	0.05
T20, 20 weeks	8.25 (0.88)	-2.71 (-5.04 to -0.39)	0.016	7.01 (0.80)	-3.17 (-5.30 to -1.04)	0.001	0.46 (-2.05 to 2.97)	0.72	0.12
T32, 32 weeks	8.00 (0.82)	-2.96 (-5.16 to -0.76)	0.004	6.60 (0.81)	-3.58 (-5.74 to -1.42)	0.0003	0.63 (-1.83 to 3.08)	0.62	0.17
GAD-7									
T0, baseline	9.63 (0.75)			8.44 (0.75)					
T4, 4 weeks	8.35 (0.87)	-1.28 (-3.60 to 1.03)	0.47	6.79 (0.84)	-1.65 (-3.89 to 0.58)	0.21	0.37 (-2.19 to 2.93)	0.77	0.10
T8, 8 weeks	7.86 (0.84)	-1.77 (-4.00 to 0.47)	0.17	6.41 (0.81)	-2.03 (-4.20 to 0.13)	0.07	0.27 (-2.21 to 2.74)	0.83	0.07
T20, 20 weeks	7.08 (0.89)	-2.55 (-4.92 to -0.19)	0.029	6.75 (0.81)	-1.70 (-3.86 to 0.47)	0.17	-0.86 (-3.40 to 1.69)	0.51	-0.23
T32, 32 weeks	6.77 (0.84)	-2.86 (-5.10 to -0.63)	0.007	6.07 (0.83)	-2.38 (-4.57 to -0.18)	0.029	-0.49 (-2.98 to 2.00)	0.70	-0.13
ISI									
T0, baseline	7.74 (0.85)			9.67 (0.85)					
T4, 4 weeks	6.41 (0.94)	-1.33 (-3.39 to 0.72)	0.33	7.50 (0.92)	-2.17 (-4.15 to -0.19)	0.027	0.84 (-1.43 to 3.11)	0.47	0.18
T8, 8 weeks	6.98 (0.91)	-0.76 (-2.75 to 1.22)	0.77	6.70 (0.89)	-2.97 (-4.89 to -1.05)	0.0007	2.21 (0.01 to 4.40)	0.049	0.46
T20, 20 weeks	6.31 (0.95)	-1.43 (-3.53 to 0.67)	0.29	6.94 (0.90)	-2.72 (-4.65 to -0.80)	0.002	1.30 (-0.97 to 3.56)	0.26	0.27
T32, 32 weeks	7.33 (0.91)	-0.41 (-2.39 to 1.57)	0.97	6.43 (0.91)	-3.23 (-5.18 to -1.28)	0.0003	2.82 (0.61 to 5.03)	0.013	0.59
EUROHIS-Qol-8									
T0, baseline	23.67 (0.73)			24.37 (0.73)					
T4, 4 weeks	26.02 (0.84)	2.36 (0.26 to 4.45)	0.022	26.77 (0.81)	2.40 (0.38 to 4.42)	0.014	-0.04 (-2.36 to 2.27)	0.97	-0.01
T8, 8 weeks	26.16 (0.81)	2.49 (0.47 to 4.52)	0.010	26.69 (0.79)	2.32 (0.36 to 4.28)	0.014	0.17 (-2.07 to 2.41)	0.47	0.05
T20, 20 weeks	26.43 (0.86)	2.77 (0.63 to 4.91)	0.006	27.98 (0.79)	3.61 (1.66 to 5.57)	<0.0001	-0.85 (-3.16 to 1.46)	0.47	-0.26
T32, 32 weeks	27.68 (0.81)	4.01 (1.99 to 6.03)	<0.0001	28.06 (0.80)	3.69 (1.70 to 5.68)	<0.0001	0.32 (-1.94 to 2.58)	0.78	0.10
FFMQ-SF									
T0, baseline	67.26 (1.77)			69.15 (1.77)					
T4, 4 weeks	68.60 (1.96)	1.34 (-2.89 to 5.57)	0.87	72.06 (1.91)	2.91 (-1.17 to 7.00)	0.24	-1.57 (-6.25 to 3.10)	0.51	-0.19
T8, 8 weeks	73.23 (1.91)	5.97 (1.89 to 10.05)	0.002	73.04 (1.87)	3.89 (-0.06 to 7.84)	0.06	2.08 (-2.44 to 6.60)	0.37	0.25
T20, 20 weeks	75.82 (1.99)	8.56 (4.24 to 12.88)	<0.0001	74.34 (1.87)	5.19 (1.23 to 9.14)	0.005	3.37 (-1.29 to 8.03)	0.16	0.40
T32, 32 weeks	75.45 (1.90)	8.19 (4.11 to 12.27)	<0.0001	74.34 (1.89)	5.19 (1.17 to 9.21)	0.006	3.00 (-1.56 to 7.56)	0.20	0.36
PSS-10									
T0, baseline	26.19 (0.97)			23.70 (0.97)					
T4, 4 weeks	22.96 (1.12)	-3.22 (-6.11 to -0.33)	0.023	21.39 (1.08)	-2.31 (-5.09 to 0.47)	0.14	-0.92 (-4.11 to 2.28)	0.57	-0.20
T8, 8 weeks	21.50 (1.08)	-4.68 (-7.47 to -1.89)	0.0002	21.56 (1.05)	-2.14 (-4.84 to 0.55)	0.16	-2.54 (-5.62 to 0.55)	0.11	-0.55
T20, 20 weeks	20.61 (1.15)	-5.57 (-8.52 to -2.62)	<0.0001	20.50 (1.05)	-3.20 (-5.90 to -0.50)	0.014	-2.37 (-5.55 to 0.81)	0.14	-0.51
T32, 32 weeks	21.30 (1.08)	-4.88 (-7.67 to -2.09)	<0.0001	20.23 (1.06)	-3.47 (-6.21 to -0.73)	0.008	-1.41 (-4.52 to 1.70)	0.37	-0.30
Resilience Scale									

(*continued on next page*)

Table 4 (*continued*)

	Intervention group not on AD (*n*= 27)			Control group not on AD (*n*= 27)					
	Mean values (SE)	Mean change (95% CI)	*P**	Mean values (SE)	Mean change (95% CI)	*P**	Mean difference (SE)	Group *Time interaction, p	Intervention effect size, Cohen's *d*
T0, baseline	79.04 (1.97)			78.15 (1.97)					
T4, 4 weeks	82.30 (2.16)	3.26 (-1.20 to 7.72)	0.23	82.22 (2.11)	4.07 (-0.24 to 8.37)	0.07	-0.81 (-5.74 to 4.13)	0.75	-0.08
T8, 8 weeks	86.51 (2.11)	7.47 (3.17 to 11.77)	0.0001	83.53 (2.07)	5.38 (1.21 to 9.55)	0.006	2.09 (-2.68 to 6.86)	0.39	0.22
T20, 20 weeks	87.02 (2.19)	7.99 (3.43 to 12.54)	<0.0001	83.81 (2.07)	5.67 (1.50 to 9.84)	0.004	2.32 (-2.59 to 7.24)	0.35	0.24
T32, 32 weeks	86.76 (2.11)	7.73 (3.43 to 12.03)	<0.0001	84.88 (2.09)	6.73 (2.50 to 10.97)	0.0005	0.99 (-3.81 to 5.80)	0.68	0.10

Abbreviations: SE, Standard Error; CI, Confidence Interval; AD, Antidepressants; PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; ISI, Insomnia Severity Index; EUROHIS-Qol-8, EUROHIS-Qol 8-item index; FFMQ-SF, Five Facet Mindfulness Questionnaire-Short Form; PSS-10, Perceived Stress Scale.

patients not on AD was greater reduction in sleep problems among controls. Our findings from these post-hoc analyses hence suggest that factors either directly or indirectly related to AD contribute to the capability of benefitting from the MHP among patients with major depression.

The improved treatment response among those with combined AD and psychological treatment is supported by earlier evidence: a meta-analysis showed that combined treatment against depression was more effective ($d = 0.35$) than the same psychological treatment alone (Cuijpers et al., 2009b). Inversely, other meta-analyses demonstrated psychotherapy ($d = 0.31$) (Cuijpers et al., 2009a) and other psychological interventions (Pampallona et al., 2004) combined with pharmacotherapy having additional value compared to pharmacotherapy alone in the treatment of depression. Furthermore, in AD-resistant clinical depression, adding smartphone CBT to medication was more effective than treatment by AD change alone (Mantani et al., 2017). Neurobiological mechanism for the improved treatment response of the combined treatment has also been demonstrated, as ADs have been shown to reactivate juvenile-like plasticity in the adult cortex in humans and animals (Castrén, 2013; Sharif et al., 2019), which may allow the enhanced influence of the psychological treatment.

There are only few previous RCTs of mobile app-interventions for depression among patient samples. Perhaps most comparable with ours, a recent trial (Graham et al., 2020) of a coach-supported platform of a suite of apps versus usual treatment among primary care patients with depression or anxiety demonstrated between group effect sizes of 0.78 and 0.64 at the completion of the 8-week intervention, respectively. The intervention effects were sustained from posttreatment to 16-week follow-up. In another RCT of the 6-month guided access to online computerized CBT versus usual care for mood and anxiety disorders in primary care (Rollman et al., 2018), between group effect sizes of 0.31 and 0.26 were reported at 6-month follow-up; these persisted 6-months later, and completing more sessions produced greater effect sizes, but combining an internet support group with computerized CBT provided no additional benefit. Subgroup results by AD were not reported in either trial. Both defined depression with PHQ-9 instead of clinical diagnosis as in the current trial. Both had college graduates without clinical experience as intervention coaches, whereas MHP has master's level mental health professionals as therapists. In a meta-analysis (Firth et al., 2017) of RCT's, depressive symptoms were reduced significantly more in smartphone apps group than in the overall control group ($g = 0.38$, $p<0.001$; comparison to inactive controls $g = 0.56$; comparison to active control treatments $g = 0.22$). The overall lost-to-follow-up rate in our RCT was 12.9%, which fits the range reported in most previous studies. A meta-analysis of smartphone-based interventions for depression symptoms reported drop-out around 30% (Firth et al., 2017). Using similar definition as in this trial, the RCT among primary care patients with depression/ anxiety in turn reported very low, 3.4%, lost-to-follow-up rate (Graham et al., 2020).

Are there comparable previous reports as related to the lack of main-group differences in depression symptom change, but improvements in resilience, experience of stress, and mindfulness skills, as in our trial? In psychotherapy research, exclusive focus on symptom reduction to evaluate therapeutic gain, and to guide the choice of the treatment, has been questioned (Messer, 2004), while broader ability to cope with potential life stressors has been emphasized equally (Blatt et al., 2000). Resembling the findings in our study, further analysis of the NIMH Treatment for Depression Collaborative Research Program indicated lack of significant differences between brief psychotherapies and clinical management (among patients on AD and not on AD) in depression symptom reduction, as measured by primary measures (Blatt et al., 2000). In the follow-up however, patients who had received brief psychotherapy reported feeling in fuller control of important aspects in their lives, and that their treatment had enabled them to develop more adaptive coping mechanisms and to deal more effectively with their experiences of depression. The authors concluded that different types of therapeutic interventions should not only aim to reduce symptoms of depression, but also vulnerability to subsequent disruptive life experiences (Blatt et al., 2000).

4.1. Limitations

Some limitations should be noted. Due to the nature of the study design, we were unable to detail TAU (appointment type/ frequency); the same was true for specific depression diagnoses and psychiatric comorbidity. We had concern over the possibility that the intervention group received significantly less TAU compared to the control group, among whom TAU was the only treatment. Due to ethical reasons, the intervention group however needed to have access to TAU. Social support was assessed informally by a study therapist instead of using validated measures. Sociodemographic data was collected scarcely and was not statistically controlled in the analyses. Blinding of the patients, therapists or researchers for patients' group status was impossible; this awareness might have biased reporting. Analyses of the intervention effects by ADs were not pre-specified in the protocol, but since the question was clinically highly relevant, we decided that it was appropriate to report these post-hoc analyses.

5. Conclusions

Therapist-guided intervention delivered through a smartphone app plus TAU was not more effective against clinical depression among young adults than TAU alone. However, examination of a subgroup of patients on AD revealed an intervention effect that appeared at the end of the 6-month follow-up. Future studies in patient samples with major depression are needed with focus on whether and how depression medication, and other potential contributing factors, modify the efficacy of smartphone app-based interventions.

Authors' contribution

A.R., R.A.H. and O.H. contributed to the design and implementation of the study. T.V. performed the statistical analyses, and A.R., T.V., V.F. H., and T.K. carried out the interpretation of the data. All authors provided critical feedback and helped shape the research, analyses and manuscript, and all accepted the final manuscript.

Funding

Study was funded by Meru Health Inc. Meru Health was involved (see competing interest, role of the authors in the company) in study design, manuscript preparation, and the decision to submit the manuscript for publication. Statistical analyses were conducted by an independent statistician (Mr. Vahlberg).

Declaration of Competing Interest

Drs. Raevuori, Hilgert, and Forman-Hoffman have equity ownership in Meru Health Inc. Drs. Hilgert and Forman-Hoffman are employed by Meru Health, and Dr. Raevuori has been employed by Meru Health. Dr. Korhonen has served as a consultant for Pfizer on nicotine dependence. Dr. Aittakumpu-Hyden's travel expenses and participation fee to the World Congress of ADHD have been covered by Shire. Mr. Vahlberg reports no financial relationships with commercial interests.

Acknowledgment

We are grateful for Drs. Pauli Tossavainen, Päivi Metsäniemi, and Tommi Väyrynen at the Finnish Student Health Service, who supported implementation of the study. We are also grateful for general practitioners, psychiatrists and other staff involved in patient recruitment at Finnish Student Health Service. We thank Kristian Ranta, Riku Lindholm, and Albert Nazander at Meru Health for their extensive efforts for the study. We thank Vilma Viiala and other staff at Meru Health who collected and archived the data. We thank intervention therapists Kaisa Koskinen and Tiina Lappalainen for their work with patients. We wish to express our gratitude to the participating patients, without whom this study would not have been possible.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2021.02.007.

References

Barkil-Oteo, A., 2013. Collaborative care for depression in primary care: how psychiatry could "troubleshoot" current treatments and practices. Yale J. Biol. Med. 86, 139.
Beck, A.T., 1976. Cognitive Therapy and the Emotional Disorders. International Universities Press, Oxford, England.
Blatt, S., Zuroff, D., Bondi, C., Sanislow, C., 2000. Short-and long-term effects of medication and psychotherapy in the brief treatment of depression: further analyses of data from the NIMH TDCRP. Psychother. Res. 10, 215–234.
Bohlmeijer, E., Ten Klooster, P.M., Fledderus, M., Veehof, M., Baer, R., 2011. Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. Assess 18, 308–320.
Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., Hedman-Lagerlöf, E., 2018. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn. Behav. Ther. 47, 1–18.
Castrén, E., 2013. Neuronal network plasticity and recovery from depression. JAMA Psychiatry 70, 983–989.
Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. J. Health Soc. Behav. 24, 385–396.
Cuijpers, P., Dekker, J., Hollon, S.D., Andersson, G., 2009a. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. J. Clin. Psychiatry 70, 1219–1229.
Cuijpers, P., van Straten, A., Warmerdam, L., Andersson, G., 2009b. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depress. Anxiety 26, 279–288.
Economides, M., Ranta, K., Nazander, A., Hilgert, O., Goldin, P.R., Raevuori, A., Forman-Hoffman, V., 2019. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR Mhealth Uhealth 7 (26), e14284.
Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., Sarris, J., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 16, 287–298.
Graham, A.K., Greene, C.J., Kwasny, M.J., Kaiser, S.M., Lieponis, P., Powell, T., Mohr, D. C., 2020. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 77, 906–914.
Jacobson S, Neil, Martell R, Christopher, Dimidjian, Sona, 2001. Behavioral activation treatment for depression: returning to contextual roots. Clin Psychol: Sci Pract 8 (3), 255–270.
Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B., 1996. Usability Evaluation in Industry. Taylor & Francis, London, pp. 189–195.
Kabat-Zinn, J., 1982. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. Gen. Hosp. Psychiatry 4, 33–47.
Kerst, A., Zielasek, J., Gaebel, W., 2020. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur. Arch. Psychiatry Clin. Neurosci. 270, 139–152.
Kohn, R., Saxena, S., Levav, I., Saraceno, B., 2004. The treatment gap in mental health care. Bull. WHO 82, 858–866.
Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
Lee, E.-H., 2012. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs. Res. 6, 121–127.
Lewis, J.R., 2018. The System Usability Scale: past, present, and future. Int. J. Hum.–Comput. Interact. 34, 577–590.
Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y., 2008. Validation and standardization of the Generalized Anxiety Disorder screener (GAD-7) in the general population. Med Care 266–274.
Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med. Care 42, 1194–1201.
Malhi, G.S., Mann, J.J., 2018. Depression. Lancet 392, 2299–2312. https://doi.org/10.1016/S0140-6736(18)31948-2.
Mantani, A., Kato, T., Furukawa, T.A., Horikoshi, M., Imai, H., Hiroe, T., Chino, B., Funayama, T., Yonemoto, N., Zhou, Q., 2017. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. JMIR 19, e373.
Messer, S.B., 2004. Evidence-based practice: beyond empirically supported treatments. Prof. Psychol. 35, 580.
Morin, C.M., Belleville, G., Bélanger, L., Ivers, H., 2011. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 34, 601–608.
Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B., Munizza, C., 2004. Combined pharmacotherapy and psychological treatment for depression: a systematic review. Arch. Gen. Psychiatry 61, 714–719.
Richards, D., 2011. Prevalence and clinical course of depression: a review. Clin. Psychol. Rev. 31, 1117–1125.
Richards, D., Richardson, T., 2012. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin. Psychiatry Rev. 32, 329–342.
Rollman, B.L., Belnap, B.H., Abebe, K.Z., Spring, M.B., Rotondi, A.J., Rothenberger, S.D., Karp, J.F., 2018. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 75, 56–64.
SAS Institute Inc 2014. Cary, NC, USA.
Schmidt, S., Mühlan, H., Power, M., 2006. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur. J. Public Health 16, 420–428.
Segal, J., Williams, J., Teasdale, J., 2002. Mindfulness-based Cognitive Therapy for Depression: a new Approach to Preventing Relapse. Guilford Press, New York, NY.
Sharif, M.H., Talebnejad, M.R., Rastegar, K., Khalili, M.R., Nowroozzadeh, M.H., 2019. Oral fluoxetine in the management of amblyopic patients aged between 10 and 40 years old: a randomized clinical trial. Eye 33, 1060–1067.
Strain, J.J., 2018. The psychobiology of stress, depression, adjustment disorders and resilience. World J. Biol. Psychiatry 19, S14–S20.
Wagnild, G., Young, H., 1993. Development and psychometric evaluation of the Resilience Scale. J. Nurs. Meas. 1, 165–178.
Williams, M.J., Dalgleish, T., Karl, A., Kuyken, W., 2014. Examining the factor structures of the five facet mindfulness questionnaire and the self-compassion scale. Psychol. Assess. 26, 407.
World Health Organization, 2020. The Impact of COVID-19 on Mental, Neurological and Substance use Services: Results of a Rapid Assessment (No. License: CC BY-NC-SA 3.0 IGO).