nature medicine

Article

Cognitive behavioral therapy skills via a smartphone app for subthreshold depression among adults in the community: the RESiLIENT randomized controlled trial

Received: 21 October 2024

Accepted: 5 March 2025

Published online: 23 April 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Subthreshold depression, defined as a depressive status falling short of the diagnostic threshold for major depression, is common, disabling and constitutes a risk factor for future depressive episodes. Cognitive behavioral therapies (CBT) have been shown to be effective but are usually provided as packages of various skills. Little research has been done to investigate whether all their components are beneficial and contributory to mental health promotion. We addressed this issue by developing a smartphone CBT app that implements five representative CBT skills (behavioral activation, cognitive restructuring, problem solving, assertion training and behavior therapy for insomnia), and conducting a master randomized study that included four 2 × 2 factorial trials to enable precise estimation of skill-specific efficacies. Between September 2022 and February 2024, we recruited 3,936 adult participants with subthreshold depression. Among those randomized, the follow-up rate was 97% at week 6 and adherence to the app was 84%. The study showed that all included CBT skills and their combinations differentially beat all three control conditions of delayed treatment, health information or self-check, with effect sizes ranging between -0.67 (95% confidence interval: -0.81 to -0.53) and -0.16 (-0.30 to -0.02) for changes in depressive symptom severity from baseline to week 6, as measured with the Patient Health Questionnaire-9 scores. Knowledge of the active ingredients of CBT can better inform the design of more effective and scalable psychotherapies in the future. (UMIN Clinical Trials Registry UMIN000047124).

Subthreshold depression, or mild depression falling short of the diagnostic threshold for major depressive disorder, is prevalent, persistent and disabling. Its prevalence is estimated to be around 11% across the world¹. It often runs a fluctuating yet chronic course^{2,3}. Subthreshold depression is associated with impaired social function and decreased quality of life^{4,5}, increased use of health services⁶ and mortality⁷. Although these negative impacts may be less grave than those associated with fully syndromal major depression at the individual level, given the sheer prevalence of those affected, subthreshold depression is responsible for economic losses to society that are indeed comparable to those of major depressive disorder⁸. Moreover, people with subthreshold depression are three times more likely to develop a major depressive episode than those without^{1,9}.

Subthreshold depression is responsive to treatments¹⁰, which can decrease the symptom burden, promote mental wellbeing and also decrease future major depressive episodes^{11,12}. Current guidelines

e-mail: furukawa@kuhp.kyoto-u.ac.jp

unanimously recommend psychotherapies, and cognitive behavioral therapies (CBT) in particular, for people with subthreshold to mild major depression^{13,14}. Given the prevalence of the condition and the human and time resources typically necessary for the face-to-face delivery of psychotherapies, digital delivery formats have received increasing attention in the past two decades¹⁵, and we now have strong evidence to show that internet CBT (iCBT) is effective for subthreshold depression¹⁶.

CBT, however, is typically provided as a package of variable combinations of cognitive and behavioral skills¹⁷, and we do not know whether all of them are contributory to mental health promotion. Such knowledge is indispensable in providing the most efficacious interventions in the most efficient way at scale in the population, because inefficacious and possibly harmful components can be dropped and more efficacious combinations can be prioritized. Efficiency and accessibility are particularly important to enhance psychological health especially among those who may feel hesitant about seeking professional help.

We therefore developed a self-help smartphone CBT app that contained modules for five representative CBT skills¹⁷: behavioral activation (BA), cognitive restructuring (CR), problem solving (PS), assertion training (AT) and behavior therapy for insomnia (BI). Briefly, BA aims to enhance mood by increasing engagement in pleasant activities, CR corrects the negative thoughts underlying depressed mood, PS teaches a structured approach to solving overwhelming problems at hand, AT helps articulate phrases to convey one's wishes without hurting others' feelings and in BI people learn and practice evidence-based efficient sleeping patterns.

We conducted a randomized controlled trial to estimate the specific efficacies of these CBT skills for adults with subthreshold depression in the general population. In this trial, dubbed the RESiLIENT trial (Resilience Enhancement with Smartphone in Living ENvironmeTs)¹⁸, we adopted a master protocol design for a randomized controlled trial because this offers the optimal design to evaluate multiple intervention components simultaneously against common controls^{19–21}. Because we may need large sample sizes to make precise estimates of the efficacy of each individual skill, we embedded four 2×2 factorial trials using each skill as a factor to make efficient use of the sample size. This paper focuses on the acute intervention effects of the five CBT skills.

Results

Participant flow and study design

Figure 1 shows the flow of participants. Of 34,123 adults assessed for eligibility, 5,364 were judged eligible, provided informed consent and were randomized to one of the 12 intervention or control arms between 5 September 2022 and 21 February 2024. Two participants later withdrew their consent, and 1,426 had baseline Patient Health Questionnaire-9 (PHQ-9) scores of \leq 4, leaving 3,936 as the intention-to-treat cohort for this study, which met the preplanned sample size requirement. Participants were randomly allocated in equal proportions to one of these 12 arms.

For the primary analysis, we embedded four 2 × 2 factorial trials. Factorial trial 1 evaluated BA and CR, trial 2 evaluated BA and PS, trial 3 evaluated BA and AT, and trial 4 evaluated BA and BI. The BA only arm and the control arm were common across the four 2 × 2 trials, thus enabling efficient use of the sample size. BA was prioritized because it was the most effective skill in the component network meta-analysis of iCBT¹⁷ and has increasingly been used in iCBT interventions aiming at scalability as BA alone or in a combination of BA plus PS¹⁶. Because there is no universally accepted gold standard control condition in psychotherapy trials^{22,23}, we included three control conditions in the master protocol: in the delayed treatment condition, participants had to wait 6 weeks before they could start active interventions; the health information arm provided information about physical exercise, nutrition and oral health on the app; and in the self-check arm participants were only prompted to fill in the PHQ-9 on the app every week. The delayed treatment was the prespecified primary control arm, but we also included the health information and self-check arms for sensitivity analyses to test whether the primary results against the delayed treatment arm also held against controls with potentially different levels of stringency.

Participant disposition

Table 1 gives the baseline characteristics of all the included participants. The participants were mainly in their 30s, 40s and 50s, 51% were men and most were married and employed. Their mean baseline PHQ-9 score was 8.1 (s.d. 2.7), 17% had suspected problematic alcohol use, 29% had a history of mental health treatment and 33% had some physical comorbidities. Supplementary Table 1 compares the distribution of these variables between the presence and absence of each intervention in the factorial design trials and shows that they were well balanced.

Primary outcome

Table 2 shows the results of the primary analyses of changes in PHQ-9 scores from baseline to week 6 (the primary outcome) for each of the four 2 × 2 factorial trials using the delayed treatment arm as the control, as prespecified in the statistical analysis plan (SAP)²⁴. In this primary analysis for each of the four factorial trials (for example, trial 1, which examined the two factors BA and CR) (Fig. 1), we estimated the efficacy of BA by comparing participants who had been allocated to that skill (for example, BA + CR or BA) with those who had not (for example, CR or delayed treatment) in a mixed-model repeated measures analysis (MMRM). The results showed that all CBT skills had specific efficacies when present compared with when the skills were absent. The standardized mean difference (SMD) was -0.38 (95% confidence interval (CI): -0.48 to -0.27, $P = 5.3 \times 10^{-13}$) for BA; $-0.27 (-0.37 \text{ to } -0.16, P = 2.9 \times 10^{-7})$ for CR; -0.27 (-0.37 to -0.17, -0.17) $P = 1.8 \times 10^{-7}$) for PS; -0.24 (-0.34 to -0.14, $P = 2.2 \times 10^{-6}$) for AT; and $-0.27 (-0.37 \text{ to } -0.16, P = 3.8 \times 10^{-7})$ for BI.

Table 3 shows the prespecified sensitivity analysis, examining the interaction of the two interventions used in each of the 2 × 2 factorial trials as hypothesized in the protocol¹⁸ and the SAP²⁴. Interactions between the two interventions were observed in all four trials. When interactions were taken account of, estimated SMDs for single-skill interventions varied from -0.52 (-0.66 to -0.38, $P = 5.9 \times 10^{-13}$) for PS to -0.65 (-0.79 to -0.51, $P = 2.2 \times 10^{-19}$) for BA, whereas those for two-skill interventions varied from -0.57 (-0.71 to -0.43, $P = 7.7 \times 10^{-15}$) for BA + BI to -0.67 (-0.81 to -0.53, $P = 2.4 \times 10^{-20}$) for BA + PS. Administering two skills did not double the efficacy of providing one skill because the interaction was antagonistic.

In the protocol¹⁸ and in the SAP²⁴ we prespecified that although research so far has not been suggestive of interactions among iCBT components^{17,25}, we would interpret the results considering the interaction if the analyses identified a strong interaction. We therefore conducted the following sensitivity analyses using more stringent control conditions of health information or self-check in each of the four 2×2 trials while including the interaction term (Extended Data Tables 1 and 2). The stronger control condition corresponded to a smaller estimated SMD. However, even when we used the most stringent control condition of self-check, we observed demonstrable specific efficacies for all the interventions, corresponding with SMDs of -0.16(-0.30 to -0.02) for PS and -0.31 (-0.45 to -0.17) for BA + PS.

The combined analysis, which conducted MMRM for the pooled dataset encompassing the four factorial trials, provides a summary of all intervention arms contrasted against all of the control arms (Table 4). In the analysis of trial 4 we excluded participants on a three-shift work schedule as per the SAP because they were considered to be less likely to benefit from BI; however, we did not exclude these participants in the combined analysis and the effect estimates for BI were essentially unchanged. The 95% CIs for the SMDs of all interventions were below zero, or the point of null effect, against all controls. The three control



Fig. 1 | **Consolidated Standards of Reporting Trials (CONSORT) diagram.** The BA arm and control arm were common across the four 2 × 2 trials and are shown by the dotted squares in trials 2 to 4. The prespecified control condition was the delayed treatment. We conducted sensitivity analyses by replacing the delayed treatment with the health information arm and the self-check arm

conditions differentiated among themselves, showing that delayed treatment was the weakest control, followed by health information and self-check in that order.

Secondary outcomes

Supplementary Tables 2–4 show the results of analyses of the main and interaction effects using MMRM for the Generalized Anxiety Disorder-7 scale (GAD-7; anxiety), the Insomnia Severity Index (ISI; insomnia) and the Short Warwick–Edinburgh Mental Well-Being Scale (SWEMWBS; mental wellbeing). Extended Data Tables 3–5 show the combined analyses for these three secondary outcomes. The health information arm was the strongest control for these outcomes.

With regard to anxiety (GAD-7), all interventions were superior to the delayed treatment or self-check controls. BA and BI were particularly efficacious for treating anxiety even when compared with the health information arm, showing SMDs of -0.24 (95% CI: -0.37 to -0.10) and -0.21 (-0.34 to -0.08), respectively. For insomnia (ISI), all interventions were superior to the delayed treatment and self-check controls, but only BI and BA + BI were superior to the health information with SMDs of -0.33 (-0.45 to -0.21) and -0.27 (-0.39 to -0.15), respectively. For mental wellbeing (SWEMWBS), almost all interventions were better than the delayed treatment control, but only BA, BI and BA + AT were better than the self-check control and none of the interventions

(Methods). ^aParticipants working on a three-shift schedule were excluded in the primary and sensitivity analyses for trial 4 (199 (5.1%) of the total sample were shift workers). AT, assertion training; BA, behavioral activation; BI, behavior therapy for insomnia; CR, cognitive restructuring; PHQ-9, Patient Health Questionnaire-9; PS, problem solving.

were superior to the health information control. The SMDs of BA, BI or BA + AT over the self-check control were 0.15 (0.03 to 0.27), 0.17 (0.05 to 0.29) and 0.13 (0.01 to 0.24).

Adherence to program and completion of follow-up assessments

Participants actively engaged with the smartphone CBT app during the initial six weeks. The chapters that constituted the app differed by week in the amount of learning that was expected of users (Extended Data Fig. 1a). Participants spent an average 20–60 min on chapter 1 (when all active intervention arm participants were introduced to a new skill) or chapter 3 (when the two-skill intervention programs began to teach the second skill); for other chapters in which reviewing was the main focus, participants tended to spend less time, between 10 and 20 min (Extended Data Table 6).

Across interventions, between 51% and 84% of participants (average 63%) completed all the intervention programs for the acute phase intervention. Between 74% and 93% of the participants for each intervention arm (average 78%) completed at least up to chapter 4 of the program covering all the major contents for the interventions (Extended Data Table 7).

The proportions of participants completing the assessments were 95.3% (3,751 of 3,936) at week 3 and 97.2% (3,824 of 3,936) at

Table 1 | Baseline demographic, psychosocial and clinical characteristics of the intention-to-treat cohort

Characteristics	Total
n	3,936
Demographic variables	
Age (mean, s.d.)	43.1 (10.7)
Gender	
Male	2,005 (51%)
Female	1,910 (49%)
Other	21 (0.5%)
Marital status	
Single	959 (24%)
Married	2,729 (69%)
Divorced or widowed	248 (6%)
Number of cohabitants including self (mean, s.d.)	2.68 (1.33)
Area Deprivation Index (score range: 0–15.9)ª (mean, s.d.)	5.63 (1.05)
Education, university or higher	2,824 (72%)
Employed	3,621 (92%)
Psychosocial variables	
Big 5 personality traits (score range: 1–5) (mean, s.d.)	
Neuroticism	3.66 (0.77)
Extraversion	2.99 (0.91)
Openness	3.15 (0.75)
Conscientiousness	3.03 (0.78)
Agreeableness	3.17 (0.85)
Adverse childhood experiences (≥4 points)	368 (9%)
Assessment of signal cases (score range: 0–3) (mean, s.d.)	
Life difficulties	0.94 (0.56)
Social support	1.31 (0.74)
Motivation	2.21 (0.39)
CBT skills (score range: 0–3) (mean, s.d.)	
ВА	1.43 (0.46)
CR	1.34 (0.59)
PS	1.75 (0.56)
AT	1.21 (0.52)
BI	1.91 (0.55)
Familiarity with smartphone (score range: $0-3$) (mean, s.d.)	1.97 (0.68)
Clinical variables	
PHQ-9 (score range: 0–27) (mean, s.d.)	8.07 (2.74)
GAD-7 (score range: 0–21) (mean, s.d.)	5.84 (3.41)
ISI (score range: 0–28) (mean, s.d.)	10.01 (4.49)
WSAS (score range: 0–8) (mean, s.d.)	1.80 (1.74)
HPQ (score range: 0–10) (mean, s.d.)	6.03 (2.06)
UWES (score range: 0–6) (mean, s.d.)	2.84 (1.28)
EQ-5D-5L utility (score range: -0.59 to 1) (mean, s.d.)	0.91 (0.09)
SWEMWBS (score range: 7–35) (mean, s.d.)	22.03 (3.68)
CAGE (≥2 points)	664 (17%)
Past or current treatments for mental health	1,160 (29%)
Hypertension	595 (15%)
Diabetes mellitus	115 (3%)

Table 1 (continued) | Baseline demographic, psychosocial and clinical characteristics of the intention-to-treat cohort

Characteristics	Total
Chronic pain	642 (16%)
Any physical disease	1,304 (33%)
Physical disability	132 (3%)

^aHigher scores indicate greater deprivation. CAGE, CAGE questionnaire; EQ-5D-5L, EuroQOL-5D-5L; HPQ, Health and Work Performance Questionnaire; UWES, Utrecht Work Engagement Scale; WSAS, Work and Social Adjustment Scale.

week 6 (Fig. 1 and Supplementary Table 5). In total, 2.8% of participants (107 of 3,824) confirmed that they had sought mental health care from a professional (psychiatrist or psychologist) by the week 6 follow-up.

Safety

As of the close date for the acute phase intervention (19 April 2024), we observed 225 warnings of increased depression and suicidality as indicated by PHQ-9 scores of \geq 10 and an item 9 (suicidality) score of \geq 2, of which 53 warnings happened on two consecutive occasions for 39 users (1% of 3,936 participants). No serious adverse events have been reported. These adverse events were too infrequent to allow meaning-ful comparisons among arms.

Sensitivity analyses

Supplementary Tables 6–9 show the results of all the prespecified sensitivity analyses subgrouping participants by baseline ISI scores to see whether the scores moderated the primary outcome of the PHQ-9, or limiting participants to those with elevated baseline scores for each secondary outcome to mitigate floor and/or ceiling effects. All the sensitivity analyses corroborated the main findings regarding the primary and secondary outcomes.

Post hoc analyses

We added two post hoc analyses to facilitate interpretation of the primary analysis results, first by extending the follow-up and second by component analysis of CBT skills.

We followed up the cohort to week 26. Participants continued to have free access to the app but, after week 6, and especially after week 10 when most of the participants had completed the program, they accessed the app for only 1–4 min per month on average (Extended Data Table 6). By week 30, the rate of adherence to the interventions increased to an average of 79% (range: 66–92%) for the whole program and 84% (range: 73–94%) for the main parts of the program (Extended Data Table 7). Supplementary Table 10a,b shows the results of analyses of the main and interaction effects, and Extended Data Table 8 shows combined analysis of the PHQ-9 up to week 26. All the interventions maintained their superiority compared with the two controls at week 26, and the two-skill interventions were generally superior to single-skill interventions with an SMD of -0.08 (-0.16 to -0.00, P = 0.049). The health information and self-check were no longer differentiated as control conditions (Extended Data Fig. 2).

Supplementary Tables 11 and 12 show the component analyses at weeks 6 and 26 respectively, with or without interaction terms in the MMRM. Table 5 juxtaposes the results with interaction terms from the two timepoints. At week 6, nonspecific treatment effects were beneficial (reducing the total effect size by -0.18 (-0.32 to -0.04)), while the waiting list effect was harmful (increasing the total effect size by 0.17 (0.03 to 0.31)). Specific efficacies of individual cognitive or behavioral skills were evident, while their interactions also tended to be influential. At week 26, however, the nonspecific effects were diminished (0.08, 95% CI: -0.09 to 0.25), and interactions among the skills also grew weaker. The specific effects of individual cognitive or behavioral skills tended to increase in size.

Trial ^a	Component	n		C	hange sco	ores from l	baseline of	9	Differences in change scores of the PHQ-9 at week 6			
				Estimate	d least squ	uares mea	n change ^b		Effect size°	Estimated least	Effect size°	P value
			Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 6	squares mean difference		
	BA ^d											
Trial 1	Present	656	-1.47	-1.98	-2.23	-2.43	-2.51	-2.89	-1.07 (-1.37 to -0.77)	-1.28 (-1.62 to -0.93)	-0.38 (-0.48 to -0.27)	5.3×10 ⁻¹³
	Absent	656	-0.70	-1.00	-1.29	-1.44	-1.69	-1.62	-0.60 (-0.90 to -0.30)	-	-	
	CR											
Trial 1	Present	657	-1.27	-1.81	-2.15	-2.29	-2.36	-2.71	-1.00 (-1.30 to -0.70)	-0.90 (-1.25 to -0.56)	-0.27 (-0.37 to -0.16)	2.9×10 ⁻⁷
	Absent	655	-0.90	-1.16	-1.37	-1.58	-1.84	-1.80	-0.67 (-0.97 to -0.37)	-	-	
	PS											
Trial 2	Present	657	-1.41	-2.05	-2.33	-2.34	-2.69	-3.08	-1.14 (-1.40 to -0.88)	-0.91 (-1.25 to -0.57)	-0.27 (-0.37 to -0.17)	1.8×10⁻ ⁷
	Absent	655	-1.35	-1.85	-1.73	-1.99	-2.00	-2.17	-0.81 (-1.07 to -0.55)	-	-	
	AT											
Trial 3	Present	656	-1.34	-2.05	-2.32	-2.58	-2.76	-3.06	-1.13 (-1.44 to -0.81)	-0.87 (-1.22 to -0.51)	-0.24 (-0.34 to -0.14)	2.2×10 ⁻⁶
	Absent	655	-1.36	-1.71	-1.76	-2.11	-2.21	-2.19	-0.81 (-1.12 to -0.50)	-	-	
	Ble											
Trial 4	Present	627	-2.10	-2.52	-2.68	-3.01	-3.08	-3.38	-1.29 (-1.58 to -0.99)	-0.94 (-1.30 to -0.58)	-0.27 (-0.37 to -0.16)	3.8×10 ⁻⁷
	Absent 624 -1.78 -1.95 -1.98 -2.21 -2.40 -2.4		-2.44	-0.93 (-1.23 to -0.64)	-	-						

Table 2 | Analysis of main effects using MMRM for PHQ-9 score at week 6, with the delayed treatment as the control arm

Least squares mean changes and differences in the PHQ-9 scores were estimated using MMRM. The regression models included fixed effects of treatment, visit (categorical) and treatment-byvisit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. *P* values are two-sided and no adjustments are made for multiple comparisons. "See Fig. 1 for the designs of trials 1 to 4. ^bA minus value indicates a reduction in depression. Values in parentheses are 95% Cls. "Effect sizes (SMD) were calculated using the observed baseline s.d. for within-group change scores and the observed week 6 s.d. for group differences. ^dThe effect size estimate of BA was obtained from the first 2×2 trial involving BA and CR, as per the SAP. The effect size estimates of BA from trials 2, 3 and 4 were -0.40 (-0.50 to -0.30), -0.35 (-0.45 to -0.25) and -0.30 (-0.40 to -0.20), respectively. [®]Participants working on a three-shift schedule were excluded in the primary and sensitivity analyses for trial 4.

Discussion

We developed a self-help smartphone CBT app involving five representative cognitive or behavioral skills (BA, CR, PS, AT and BI) and conducted an individually randomized controlled trial of these interventions for subthreshold depression^{26,27}. We achieved a follow-up rate of 97% at week 6 and the average completion rate for the program was 84%. All skills and their combinations were differentially superior to the delayed treatment, health information and self-check controls, with effect sizes ranging between 0.67 and 0.16 for the primary outcome of depression at week 6. These interventions were also differentially efficacious for the secondary outcomes of anxiety, insomnia and wellbeing. The efficacy of cognitive and behavioral interventions was maintained up to week 26.

In a previous component network meta-analysis of iCBT, we found that BA was the most efficacious component; there was some suggestive evidence supporting BI, PS and AT, but no strong support for CR¹⁷. The results of the current study are aligned with these estimates because the 95% CIs of the effect sizes of the components in our study overlapped with those from the meta-analysis. However, our study demonstrated specific efficacies of individual cognitive or behavioral skills directly in a single randomized trial and yielded more precise effect estimates for these components because of its large sample size, fixed intervention components and master protocol design. Moreover, we were able to demonstrate the efficacy of CR per se, in contrast to the meta-analysis.

We prioritized BA in the design of the four 2×2 factorial trials because previous studies have shown it to be among the more efficacious of the skills¹⁷ and therefore a good candidate for inclusion in future iCBT packages¹⁶ or even as a standalone intervention²⁸. As expected, BA was among the more efficacious skills at week 6 (Table 4), but seemed to lose its edge at week 26 (Extended Data Table 8). There was greater adherence to BA than to the other skills, except for BI (Extended Data Table 7). However, combining BA with another skill did not result in a straightforward sum of the single-skill intervention arms at week 6. In retrospect, this was to be expected because all intervention arms must have had some common nonspecific treatment effects due to the placebo effect, self-monitoring and encouragement emails. Intervention arms teaching two skills could not double these effects and therefore were not a simple sum of the efficacies of single-skill interventions. This was evident in analyses of the interaction terms of factorial trials (Table 3) and was corroborated by the component analyses at week 6 (Table 5). When we followed up the cohort to week 26, however, there were no longer demonstrable nonspecific treatment effects and interactions among the skills became less influential. Individual skills maintained their effect sizes. In other words, learning skills continued to be beneficial up to week 26, and learning two skills was more beneficial than learning one.

The five cognitive or behavioral skills were also differentially efficacious for the secondary outcomes of anxiety, insomnia and mental

Trial^a Arm Change scores from baseline of the PHQ-9 Differences in change scores of the PHQ-9 at week 6 n Estimated least squares mean change^b Effect size^c Estimated least P value P value Effect size squares mean (main (interaction Week 1 Week 2 Week 3 Week 4 Week 5 Week 6 Week 6 difference effect) of two components)^c BA+CR329 -1.40 -1.97 -2.14 -2.36 -2.32 -2.82 -1.04 -2.17 -0.64 2.1×10⁻¹⁸ 6 2×10⁻⁸ (-0.78 to -0.50) (-1.35 to -0.73) (-2.65 to -1.69) 4.4×10⁻¹⁹ ΒA 327 -144 -1.88 -2.21 -2.40 -2.60 -2.86 -1.06 -2.22 -0.65 (-1.37 to -0.75) (-2.70 to -1.74) (-0.79 to -0.51) Trial 1 CR -1.84 -0.54 8.9×10⁻¹⁴ 328 -1.03 -1.56-2.04 -2.12 -2.29 -2.48 -0.92 (-1.23 to -0.61) (-2.32 to -1.36) (-0.68 to -0.40) Delaved 328 -0.42 -0.65 -0.24 treatment (-0.55 to 0.07) (reference) -1.24 -2.27 2.4×10⁻²⁰ 6.6×10⁻⁷ BA+PS -1.59 -2.11 -2.70 -2.56-3.03 -3.34 -0.67 329 (-1.51 to -0.97) (-2.74 to -1.80) (-0.81 to -0.53) ΒA 327 -1.86 -2.31 -2.63 -2.83 -3.03 -3.29 -1.22 -2.22 -0.65 2.2 × 10⁻¹⁹ (-1.50 to -0.95) (-2.70 to -1.75) (-0.79 to -0.51) Trial 2 PS 5.9 × 10⁻¹³ 328 -1.24 -2.00 -1.97 -2.14 -2.35 -2.84 -1.05 -1.77 -0.52 (-1.33 to -0.78) (-2.24 to -1.29) (-0.66 to -0.38) -0.40 Delaved 328 -0.85 -1.07 treatment (-0.67 to -0.12) (reference) BA+AT 328 -1.42 -2.04 -2.44 -2.60 -2.77 -3.17 -1.17 -2.08 -0.59 5.1×10^{-16} 3.8×10⁻⁸ (-1.49 to -0.85) (-2.58 to -1.59) (-0.73 to -0.45) ΒA 327 -1.88 -2.33 -2.65 -2.85 -3.05 -3.31 -1.22 -2.22 -0.63 6.4 × 10⁻¹⁸ (-1.54 to -0.90) (-2.72 to -1.73) (-0.77 to -0.49) Trial 3 AT 4.6 × 10⁻¹³ 328 -1.26 -2.05 -2.20 -2.55 -2.74 -2.94 -1.08 -1.85 -0.52 (-1.41 to -0.76) (-2.35 to -1.35) (-0.66 to -0.38) -040 Delaved 328 _ -0.87 -109 (-0.72 to -0.08) treatment (reference) 3.8×10⁻¹¹ BA+BI 311 -1.78 -2.23 -2.38 -2.84 -2.84 -3.22 -1.23 -2.00 -0.57 7.7 × 10⁻¹⁵ (-1.54 to -0.92) (-2.50 to -1.50) (-0.71 to -0.43) ΒA 1.9 × 10⁻¹⁸ 314 -2.06 -2.48 -2.79 -3.01 -3.21 -3.49 -2.27 -0.64 -1.33 (-1.64 to -1.03) (-2.77 to -1.77) (-0.79 to -0.50) Trial 4^e BI -0.60 1.3 × 10⁻¹⁶ 313 -2.23-2.61-2.79-3.00 -3.13 -3.35 -1.28-213(-1.58 to -0.98) (-2.63 to -1.63) (-0.75 to -0.46) Delayed 313 -0.98 -1.22 -0.47 treatment (-0.77 to -0.16) (reference)

Table 3 | Analysis of main and interaction effects using MMRM for PHQ-9 score at week 6, with the delayed treatment as the control arm

Least squares mean changes and differences in the PHQ-9 scores were estimated using MMRM. The regression models included fixed effects of treatment, visit (as categorical) and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. The interaction *P* values are calculated from the MMRM involving interaction terms for the corresponding treatments. *P* values are two-sided and no adjustments are made for multiple comparisons. "See Fig. 1 for the design of trials 1 to 4. ^bA minus value indicates a reduction in depression. Values in parentheses are 95% CIs. "Effect sizes (SMD) were calculated using the observed baseline s.d. for within-group change scores and the observed week 6 s.d. for group differences. ⁴A sensitivity analyses with linear regression analyses using only week 6 data for each of the 2×2 trials confirmed that all interaction terms had *P* values <0.001. ^eParticipants working on a three-shift schedule were excluded in the primary and sensitivity analyses for trial 4.

wellbeing. The intervention arms BA, BI and probably to a lesser degree BA + CR or BA + PS were particularly efficacious in reducing anxiety symptoms. As expected, BI and BA + BI had the largest effect sizes in improving sleep. For positive mental wellbeing, BA, BI and BA + AT were more promotional than the delayed treatment or self-check controls. Although all the cognitive or behavioral skills we included in the app were beneficial overall, these differentiations provide the first step in matching interventions with the status and needs of individual participants. A future study should also take into account participants' other characteristics to personalize and optimize the interventions.

We were able to monitor participant access to the program through the remote server. The access logs indicated that the participants' app usage occurred in short, intermittent sessions, typically lasting between 2 and 10 min. This suggests that participants engaged with the app during brief moments of free time throughout their day. The amount of time spent accessing the app dramatically decreased after week 6 or week 10, by which time most participants had finished all of the lessons in their allocated intervention. However, the efficacies of the CBT skills persisted over the 26 weeks of follow-up. We interpret this as a sign of true learning: once users master the essentials of BA, CR or any other CBT skill, they do not need to go back to the app but can practice them on their own and continue to benefit from the learned skills.

The rate of adherence to the program, defined as completing the main learning materials for the interventions, reached 78% by week 10 and 84% by week 30, on average. These figures compare favorably with those reported in the literature for similarly defined adherence rates of iCBT programs, which are 30% for unguided iCBT²⁹ or 68% for therapist-guided iCBT³⁰. Throughout the acute intervention and follow-up, PS, CR, BA + PS and BA + CR tended to have lower adherence

Table 4 | Combined analysis using the MMRM for PHQ-9 score at week 6

Arm	Estim	nated leas	t squares i	nean char	nge scores	s of the	Effect size for	Estimated diffe	erences in	Effect size for di	fferences in change	scores ^b (95% CI)	
				РН	Q-9°			baseline – week 6 change scores ^b	change scores Delayed Trea	s (versus atment)			
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	(95% CI)	(95% CI)	P value	versus Delayed Treatment	versus Health Information	versus Self-Check
ВА	327	-2.03	-2.48	-2.80	-3.00	-3.19	-3.46	-1.26 (-1.45 to -1.08)	-2.21 (-2.71 to -1.72)	6.2×10 ⁻¹⁸	-0.63 (-0.77 to -0.49)	-0.48 (-0.62 to -0.33)	-0.29 (-0.44 to -0.15)
CR	328	-1.63	-2.16	-2.64	-2.72	-2.89	-3.08	-1.13 (-1.31 to -0.94)	-1.83 (-2.33 to -1.34)	3.6×10 ⁻¹³	-0.52 (-0.66 to -0.38)	-0.37 (-0.51 to -0.22)	-0.18 (-0.33 to -0.04)
PS	328	-1.41	-2.18	-2.15	-2.32	-2.53	-3.01	-1.10 (-1.29 to -0.91)	-1.76 (-2.26 to -1.27)	3.3×10 ⁻¹²	-0.50 (-0.64 to -0.36)	-0.35 (-0.49 to -0.20)	-0.16 (-0.31 to -0.02)
AT	328	-1.42	-2.22	-2.35	-2.71	-2.90	-3.10	-1.13 (-1.32 to -0.95)	-1.85 (-2.34 to -1.36)	2.2×10 ⁻¹³	-0.53 (-0.67 to -0.39)	-0.37 (-0.52 to -0.23)	-0.19 (-0.33 to -0.04)
BI	328	-2.27	-2.61	-2.81	-3.06	-3.15	-3.40	-1.24 (-1.43 to -1.06)	-2.15 (-2.65 to -1.66)	1.3×10 ⁻¹⁷	-0.61 (-0.75 to -0.47)	-0.46 (-0.6 to -0.32)	-0.28 (-0.42 to -0.13)
BA+CR	329	-2.01	-2.57	-2.75	-2.97	-2.94	-3.44	-1.25 (-1.44 to -1.07)	-2.19 (-2.68 to -1.69)	6.0×10 ⁻¹⁸	-0.62 (-0.76 to -0.48)	-0.47 (-0.61 to -0.32)	-0.29 (-0.43 to -0.14)
BA+PS	329	-1.77	-2.29	-2.89	-2.74	-3.21	-3.52	-1.29 (-1.47 to -1.10)	-2.27 (-2.77 to -1.78)	2.0×10 ⁻¹⁹	-0.65 (-0.79 to -0.51)	-0.49 (-0.64 to -0.35)	-0.31 (-0.46 to -0.17)
BA+AT	328	-1.59	-2.21	-2.61	-2.76	-2.94	-3.34	-1.22 (-1.41 to -1.03)	-2.09 (-2.58 to -1.60)	1.5×10 ⁻¹⁶	-0.60 (-0.74 to -0.45)	-0.44 (-0.59 to -0.30)	-0.26 (-0.40 to -0.11)
BA+BI	327	-1.78	-2.26	-2.40	-2.82	-2.88	-3.22	-1.18 (-1.36 to -0.99)	-1.97 (-2.47 to -1.48)	5.6×10 ⁻¹⁵	-0.56 (-0.70 to -0.42)	-0.41 (-0.55 to -0.26)	-0.22 (-0.37 to -0.08)
Self-check (reference)	328	-1.37	-2.00	-2.16	-2.06	-2.50	-2.46	-0.90 (-1.08 to -0.72)	-1.21 (-1.70 to -0.72)	1.4×10 ⁻⁶	-0.35 (-0.49 to -0.21)	-0.18 (-0.33 to -0.04)	-
Health information (reference)	328	-	-	-1.61	-	-	-1.83	-0.67 (-0.85 to -0.49)	-0.58 (-1.08 to -0.09)	0.020	-0.17 (-0.31 to -0.03)	_	0.18 (0.04 to 0.33)
Delayed treatment (reference)	328	-	-	-1.03	-	-	-1.25	-0.46 (-0.64 to -0.27)	-	-	-	0.17 (0.03 to 0.32)	0.35 (0.21 to 0.50)

Least squares mean changes and differences in the PHQ-9 scores were estimated with MMRM. The regression model included fixed effects of treatment, visit (as categorical) and treatment-byvisit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. P values are two-sided and no adjustments are made for multiple comparisons. ^aA minus value indicates a reduction in depression. Values in parentheses are 95% CIs. ^bEffect sizes (SMD) were calculated using the observed baseline s.d. for within-group change scores and the observed week 6 s.d. for group differences.

rates than the other intervention arms. However, although PS tended to show smaller effect sizes, BA + PS and BA + CR were usually associated with the largest effect sizes among all the interventions.

Our results provide insights into control conditions in psychotherapy trials²². There have been several network meta-analyses suggesting that waiting list controls may be harmful and produce larger effect size estimates than no treatment for the same intervention^{23,31}. However, the comparisons were mostly indirect because few studies directly compared the waiting list against no treatment. The current study provides evidence that the waiting list is inferior to other commonly used control conditions in psychotherapy trials. The health information arm in our study, without weekly self-checks or encouragement emails from the central office, represents an 'informed, health-conscious life as usual' and may be considered the most appropriate control when the research question of the study is to ask whether it is meaningful to do something more than this informed life as usual. The self-check arm included repeated administration of the PHQ-9 and weekly encouragement email from the office; although none of these constitute specifically cognitive or behavioral interventions, they are integral parts of any guided CBT intervention and comparison with them is meaningful only when we are interested in the specific efficacies of particular intervention components (as in our component analyses). However, the self-check control would be inappropriate if the relevant study question was the efficacy of the total program including specific as well as nonspecific, but essential, ingredients of CBT. Although the self-check was the strongest control for the PHQ-9, this was not the case for the other outcomes, for which the health information was strongest. In other words, repeating self-reports may improve scores on the self-report scale, but not on others. In comparison with these, the waiting list overestimates the efficacy of the intervention by SMDs of -0.35 (-0.50 to -0.21) or -0.18 (-0.33 to -0.04) respectively, and should not be used in confirmatory trials of interventions (late phase 2 or phase 3 studies). Its use may be justifiable in exploratory early phase 2 or pragmatic phase 4 studies but in such instances interpretation of the obtained effect sizes must be nuanced accordingly. A corollary to this is that no future meta-analysis of psychotherapies should combine these different control conditions, otherwise the resultant pooled effect estimate of a therapy would be uninterpretable.

Our results contrast with the so-called Dodo bird verdict, which claims that all bona fide psychotherapies have comparable efficacies³². If this claim held true for iCBT, we should expect: (1) all skills to be equally effective for depression, anxiety, insomnia and wellbeing; and (2) one-skill interventions to match two-skill interventions in efficacy. However, we found that different cognitive or behavioral skills had varying efficacy for depression and other symptoms. Although the benefits were not additive at week 6, by week 26, learning two skills proved more efficacious than learning one skill. The inability to differentiate between psychotherapies in past research is probably due to the fact that different skills were combined in various ways and in variable quality, and also to the small sample sizes in dismantling studies (typically only ten per arm)³³. Mixing various control conditions may have further diluted the differences. Our study used standardized and well-controlled interventions to provide support for differential efficacies among the five cognitive or behavioral skills, distinct from nonspecific common effects or the nocebo effect due to the waiting list condition. Given the results from several component

Table 5 | Component analyses of individual CBT skills on the PHQ-9 scores at week 6 and week 26, using MMRM with interaction terms

Component ^a	<i>n</i> at baseline	n at week 6	Differences in ch	ange scores from b week 6	aseline to	<i>n</i> at week 26	Differences in change scores from baseline to week 26				
			Estimated least squares mean difference	Effect size°	P value	-	Estimated least squares mean difference	Effect size ^c	<i>P</i> value		
ns	3,280	3,177	-0.63 (-1.12 to -0.13)	-0.18 (-0.32 to -0.04)	0.013	2,943	0.29 (-0.30 to 0.88)	0.08 (-0.09 to 0.25)	0.330		
ba	1,640	1,581	-1.00 (-1.49 to -0.50)	-0.28 (-0.43 to -0.14)	7.7×10⁻⁵	1,452	-0.96 (-1.56 to -0.37)	-0.28 (-0.45 to -0.11)	0.002		
cr	657	633	-0.62 (-1.12 to -0.13)	-0.18 (-0.32 to -0.04)	0.013	587	-0.93 (-1.53 to -0.34)	-0.27 (-0.44 to -0.10)	0.002		
ps	657	632	-0.55 (-1.05 to -0.06)	-0.16 (-0.3 to -0.02)	0.029	582	-0.78 (-1.38 to -0.18)	-0.22 (-0.4 to -0.05)	0.011		
at	656	635	-0.64 (-1.13 to -0.15)	-0.18 (-0.32 to -0.04)	0.011	578	-0.82 (-1.42 to -0.23)	-0.24 (-0.41 to -0.06)	0.007		
bi	655	639	-0.94 (-1.44 to -0.45)	-0.27 (-0.41 to -0.13)	1.7×10 ⁻⁴	577	-1.19 (-1.79 to -0.59)	-0.34 (-0.51 to -0.17)	9.6×10 ⁻⁵		
ba×cr	-	-	0.65 (-0.05 to 1.35)	0.18 (-0.02 to 0.38)	0.070	-	0.51 (-0.34 to 1.36)	0.15 (-0.10 to 0.39)	0.237		
ba×ps	-	-	0.49 (-0.21 to 1.19)	0.14 (-0.06 to 0.34)	0.172	-	0.39 (-0.46 to 1.24)	0.11 (-0.13 to 0.36)	0.366		
ba×at	-	-	0.76 (0.06 to 1.46)	0.22 (0.02 to 0.42)	0.033	-	0.57 (-0.28 to 1.42)	0.16 (-0.08 to 0.41)	0.188		
ba×bi	-	-	1.18 (0.48 to 1.88)	0.34 (0.14 to 0.54)	0.001	-	1.21 (0.36 to 2.05)	0.35 (0.10 to 0.59)	0.005		
wl	328	325	0.58 (0.09 to 1.08)	0.17 (0.03 to 0.31)	0.020	-	-	-	-		

Least squares mean changes and differences in the PHQ-9 scores were estimated with MMRM. The regression model included fixed effects of components, their interactions with *ba*, visit (as categorical) and component-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. *P* values are two-sided and no adjustments are made for multiple comparisons. "We conceptualized BA as consisting of *ns+ba*, CR as consisting of *ns+cr*, PS as consisting of *ns+ps*, AT as consisting of *ns+at* and BI as consisting of *ns+bi*, while we regarded BA+CR as consisting of *ns+ba+cr*, BA+PS of *ns+ba+ps*, BA+AT of *ns+ba+at* and BA+BI of *ns+ba+at*. The self-check arm was conceptualized as consisting of *ns* only, as participants received encouragement emails and were asked to complete the PHQ-9 weekly. The health information arm was used to represent the natural course and the delayed treatment was conceptualized as consisting of *ns*-*bi*, *sp*-cific effect of behavioral activation'; *cr* 'specific effect of cognitive restructuring'; *ps* 'specific effect of problem solving'; at 'specific effect of sasertion training'; *bi* 'specific effect of behaviour therapy for insomnia'; and *W* 'waiting list'. ^bA minus value indicates a reduction in depression. Values in parentheses are 95% Cls. ^cEffect sizes (SMDs) were calculated using the observed week 6 s.d. for group differences.

network meta-analyses distinguishing helpful, neutral or even harmful elements in interventions traditionally classified under one name (for example, CBT for panic disorder³⁴, iCBT for depression¹⁷ and CBT for insomnia³⁵), more rigorous and granular research on the Dodo bird verdict is warranted.

Psychotherapies are complex interventions and are intrinsically multifarious. Allegedly new interventions appear every year but their efficacies per se, let alone their relative efficacies against older therapies, often remain elusive. When multiple interventions and control conditions are available, the master protocol design allows for the comparison of several interventions against common controls and the platform trial allows for the entry of new interventions (and the dropping of old, inefficacious ones)^{20,21}. Our RESiLIENT trial is one of the earlier adoptions of this advanced trial methodology in the field of psychotherapy^{19,20}. It must be pointed out that the master randomized trial not only enables efficient use of the sample size by sharing common controls, but also saves time, effort and money on the part of researchers because multiple hypotheses can be tested on the same platform, obviating the costs of starting and closing individual trials. When a new arm is added to an existing platform, recruitment can start at full speed.

This study is not without limitations. First, given the nature of the interventions, it was not possible to keep participants blind to the intervention, as is the case with all psychotherapy trials. This could lead to performance bias. However, in the RESiLIENT trial there were 12 arms and the nature of each arm—whether it was intended as an

Nature Medicine | Volume 31 | June 2025 | 1830-1839

active intervention or a control and comparison intervention-was not emphasized to participants (except for the delayed treatment where it was obvious that the active intervention was withheld to week 6). Thus participants experienced the nine intervention arms as well as the weekly self-check arm and the health information arm as if they were receiving intervention outside the trial setting. There was no apparent deviation in the interventions because of the trial setting, which should therefore not constitute a risk of bias due to deviations from intended interventions as per the new Cochrane Collaboration's Risk of Bias tool³⁶. In addition, the lack of participant blinding could also be associated with assessment bias, because all outcomes in this study were self-rated by participants. However, there is growing empirical evidence to show that self-ratings tend to lead to more conservative effect size estimates than blinded observer ratings in clinical trials³⁷⁻⁴⁰. Second, this study focused on the acute intervention effects and examined their durability up to 26 weeks. Longer follow-up and an examination of the enduring effects of the skills are warranted. Third, we examined five representative cognitive or behavioral skills but did not include relaxation, mindfulness or acceptance, among others, which sometimes are considered effective elements in broadly conceived CBT⁴¹. We limited ourselves to CBT skills based on classical behavioral or cognitive techniques that had some empirical support¹⁷, but did not include relaxation because there is evidence suggestive of its harmful effects^{17,34,35}. We did not include third-wave components such as mindfulness or acceptance because theoretically they are posited to be antithetical to CR and might present some interpretational difficulties

when combined with the selected skills. We need comparative and combinational studies for these and other new interventions before we can be certain of their inclusion in empirically proven psychotherapy arsenals. Fourth, as we accumulate knowledge about effective ingredients of psychotherapies, they may or may not be additive if they are provided altogether. In this study, two-skill interventions including the BA were not additive at week 6, but were largely so at week 26 (except for BA and Bl). Ultimately, this is another empirical question for future studies. Lastly, it is unclear whether our findings about the five skills are limited to our smartphone app only for adults with subthreshold depression and may not be generalizable to CBT skills administered using other apps or by human therapists; extrapolate to the school setting, to elderly people or to people with comorbidities; apply to populations other than Japanese; or extrapolate to major depressive disorder. These are all empirical questions that require further research.

In conclusion, CBT skills as implemented in our smartphone app proved differentially efficacious against depressive symptoms for up to 26 weeks in adults with subthreshold depression. Differential efficacies will enable efficient packaging of skills that match individual participant's characteristics and needs by providing an evidence base to decide which skills to prioritize. Careful attention must be given to the combinations of skills in the package to address potential interactions. Given the substantial and persistent global burden of depression, the widespread implementation of a scalable service platform based on this app to promote mental health among people with subthreshold depression is both timely and crucial. Leveraging smartphone technology and remote server connectivity offers a promising avenue for the continued improvement of digital interventions.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-025-03639-1.

References

- 1. Zhang, R. et al. The prevalence and risk of developing major depression among individuals with subthreshold depression in the general population. *Psychol. Med.* **53**, 3611–3620 (2023).
- Jinnin, R. et al. Detailed course of depressive symptoms and risk for developing depression in late adolescents with subthreshold depression: cohort study. *Neuropsychiatr. Dis. Treat.* 13, 25–33 (2017).
- Sigstrom, R., Waern, M., Gudmundsson, P., Skoog, I. & Ostling, S. Depressive spectrum states in a population-based cohort of 70-year olds followed over 9 years. *Int. J. Geriatr. Psychiatry* 33, 1028–1037 (2018).
- Karsten, J., Penninx, B. W., Verboom, C. E., Nolen, W. A. & Hartman, C. A. Course and risk factors of functional impairment in subthreshold depression and anxiety. *Depress. Anxiety* **30**, 386–394 (2013).
- Backenstrass, M. et al. A comparative study of nonspecific depressive symptoms and minor depression regarding functional impairment and associated characteristics in primary care. *Compr. Psychiatry* 47, 35–41 (2006).
- 6. Goldney, R. D., Fisher, L. J., Dal Grande, E. & Taylor, A. W. Subsyndromal depression: prevalence, use of health services and quality of life in an Australian population. *Soc. Psychiatry Psychiatr. Epidemiol.* **39**, 293–298 (2004).
- Cuijpers, P. et al. Differential mortality rates in major and subthreshold depression: meta-analysis of studies that measured both. Br. J. Psychiatry 202, 22–27 (2013).
- 8. Cuijpers, P. et al. Economic costs of minor depression: a populationbased study. *Acta Psychiatr. Scand.* **115**, 229–236 (2007).

- Lee, Y. Y. et al. The risk of developing major depression among individuals with subthreshold depression: a systematic review and meta-analysis of longitudinal cohort studies. *Psychol. Med.* 49, 92–102 (2019).
- 10. Cuijpers, P. et al. Psychotherapy for subclinical depression: meta-analysis. *Br. J. Psychiatry* **205**, 268–274 (2014).
- 11. Buntrock, C. et al. Psychological interventions to prevent the onset of major depression in adults: a systematic review and individual participant data meta-analysis. *Lancet Psychiatry* **11**, 990–1001 (2024).
- 12. Harrer, M. et al. Psychological intervention in individuals with subthreshold depression: individual participant data meta-analysis of treatment effects and moderators. *Br. J. Psychiatry* (in press).
- 13. Simon, G. E., Moise, N. & Mohr, D. C. Management of depression in adults: a review. *JAMA* **332**, 141–152 (2024).
- Herrman, H. et al. Time for united action on depression: a Lancet–World Psychiatric Association Commission. *Lancet* **399**, 957–1022 (2022).
- Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A. & Furukawa, T. A. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry* **76**, 700–707 (2019).
- 16. Karyotaki, E. et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* **78**, 361–371 (2021).
- 17. Furukawa, T. A. et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. *Lancet Psychiatry* **8**, 500–511 (2021).
- Furukawa, T. A. et al. Four 2×2 factorial trials of smartphone CBT to reduce subthreshold depression and to prevent new depressive episodes among adults in the community-RESiLIENT trial (Resilience Enhancement with Smartphone in Llving ENvironmenTs): a master protocol. *BMJ Open* 13, e067850 (2023).
- U.S. Food and Drug Administration. Master Protocols for Drugs and Biological Product Development: Guidance for Industry (Draft Guidance) (2023); https://www.fda.gov/ regulatory-information/search-fda-guidance-documents/ master-protocols-drug-and-biological-product-development
- 20. Gold, S. M. et al. Transforming the evidence landscape in mental health with platform trials. *Nat. Mental Health* https://doi.org/10.1038/s44220-025-00391-w (2025).
- 21. Park, J. J. H. et al. An overview of platform trials with a checklist for clinical readers. J. Clin. Epidemiol. **125**, 1–8 (2020).
- 22. Gold, S. M. et al. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *Lancet Psychiatry* **4**, 725–732 (2017).
- Michopoulos, I. et al. Different control conditions can produce different effect estimates in psychotherapy trials for depression. J. Clin. Epidemiol. 132, 59–70 (2021).
- 24. Noma, H. et al. RESiLIENT (Resilience Enhancement with Smartphone in LIving ENvironmenTs) Trial: The Statistical Analysis Plan. Preprint at https://doi.org/10.1101/2024.03.16.24304261 (2024).
- 25. Sakata, M. et al. Components of smartphone cognitive-behavioural therapy for subthreshold depression among 1093 university students: a factorial trial. *Evid. Based Ment. Health* **25**, e18–e25 (2022).
- 26. METAPSY. Meta-Analytic Database of Psychotherapy Trials (2024); https://www.metapsy.org/
- Cuijpers, P., Harrer, M., Miguel, C., Ciharova, M. & Karyotaki, E. Five decades of research on psychological treatments of depression: A historical and meta-analytic overview. *Am. Psychol.* https://doi.org/10.1037/amp0001250 (2023).

- Cuijpers, P. et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry* 20, 283–293 (2021).
- 29. Karyotaki, E. et al. Predictors of treatment dropout in selfguided web-based interventions for depression: an 'individual patient data' meta-analysis. *Psychol. Med.* **45**, 2717–2726 (2015).
- van Ballegooijen, W. et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. *PLoS ONE* 9, e100674 (2014).
- Furukawa, T. A. et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr. Scand.* 130, 181–192 (2014).
- Wampold, B. E. et al. A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, 'all must have prizes'. *Psychol. Bull.* **122**, 203–215 (1997).
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M. & Hollon, S. D. Component studies of psychological treatments of adult depression: a systematic review and meta-analysis. *Psychother. Res.* 29, 15–29 (2019).
- 34. Pompoli, A. et al. Dismantling cognitive-behaviour therapy for panic disorder: a systematic review and component network meta-analysis. *Psychol. Med.* **48**, 1945–1953 (2018).
- Furukawa, Y. et al. Components and delivery formats of cognitive behavioral therapy for chronic insomnia in adults: a systematic review and component network meta-analysis. *JAMA Psychiatry* 81, 357–365 (2024).
- 36. Sterne, J. A. C. et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* **366**, l4898 (2019).
- Cuijpers, P., Li, J., Hofmann, S. G. & Andersson, G. Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. *Clin. Psychol. Rev.* **30**, 768–778 (2010).

- Sayer, N. et al. The relations between observer-rating and self-report of depressive symptomatology. *Psychol. Assess.* 5, 350–360 (1993).
- Greenberg, R. P., Bornstein, R. F., Greenberg, M. D. & Fisher, S. A meta-analysis of antidepressant outcome under 'blinder' conditions. J. Consult. Clin. Psychol. 60, 664–669 (1992); discussion 670–667.
- 40. Miguel, C. et al. Self-reports vs clinician ratings of efficacies of psychotherapies for depression: a meta-analysis of randomized trials. *Epidemiol. Psychiatr. Sci.* **34**, e15 (2025).
- Watkins, E., Newbold, A., Tester-Jones, M., Collins, L. M. & Mostazir, M. Investigation of active ingredients within internetdelivered cognitive behavioral therapy for depression: a randomized optimization trial. *JAMA Psychiatry* 80, 942–951 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

Toshi A. Furukawa ©¹⊠, Aran Tajika², Rie Toyomoto ©², Masatsugu Sakata ©³, Yan Luo⁴, Masaru Horikoshi⁵, Tatsuo Akechi ©⁶, Norito Kawakami ©^{7,8}, Takeo Nakayama⁹, Naoki Kondo¹⁰, Shingo Fukuma^{11,12}, Ronald C. Kessler ©¹³, Helen Christensen ©^{14,15}, Alexis Whitton¹⁵, Inbal Nahum-Shani ©¹⁶, Wolfgang Lutz ©¹⁷, Pim Cuijpers ©^{18,19}, James M. S. Wason ©²⁰ & Hisashi Noma²¹

¹Kyoto University Office of Institutional Advancement and Communications, Kyoto, Japan. ²Department of Health Promotion and Human Behavior, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan. ³Department of Neurodevelopmental Disorders, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. ⁴Center for Medical Education and Internationalization, Kyoto University Graduate School of Medicine, Kyoto, Japan. ⁵Musashino University, Tokyo, Japan. ⁶Department of Psychiatry and Cognitive-Behavioral Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. ⁷Department of Digital Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ⁸The Junpukai Foundation, Okayama, Japan. ⁹Department of Health Informatics, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁰Department of Social Epidemiology, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹¹Human Health Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹²Department of Epidemiology Infectious Disease Control and Prevention, Hiroshima University Graduate School of Biomedical and Health Sciences, Hiroshima, Japan. ¹³Department of Health Care Policy, Harvard Medical School, Boston, MA, USA. ¹⁴Discipline of Psychiatry and Mental Health, University of New South Wales, Sydney, New South Wales, Australia. ¹⁵Black Dog Institute, University of New South Wales, Sydney, New South Wales, Australia. ¹⁶Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. ¹⁷Department of Psychology, Trier University, Trier, Germany. ¹⁸Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ¹⁹Babeş-Bolyai University, International Institute for Psychotherapy, Cluj-Napoca, Romania. ²⁰Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. ²¹Department of

Methods

The RESiLIENT trial is a master protocol trial involving four 2×2 factorial trials. The study was approved by Kyoto University Ethics Committee (C1556) on 3 March 2022. Recruitment of participants took place between 5 September 2022 and 21 February 2024. We preregistered the trial in the UMIN Clinical Trials Registry (UMIN000047124) on 1 August 2022. We have previously published the study protocol¹⁸. We published the SAP²⁴ on 18 March 2024 after the last-participant-in but before the last-participant-out. The protocol and SAP are available in the Supplementary Information and are summarized here. This report follows the CONSORT 2010 statement⁴² and its extensions for nonpharmacologic treatments⁴³ and for factorial randomized trials⁴⁴.

The study protocol is available as an open-access publication¹⁸ and the SAP is available on MedRxiv²⁴; both are given in the Supplementary Information.

Participants

Eligibility criteria. Inclusion criteria were as follows: (1) people of any gender, aged 18 years or older at the time of providing informed consent; (2) in possession of their own smartphone (either iOS or Android); (3) written informed consent for participation in the trial; (4) completion of all the baseline questionnaires within 1 week after providing informed consent; (5) screening PHQ-9⁴⁵ total scores of (a) between 5 and 9, inclusive, or (b) between 10 and 14, inclusive, but not scoring 2 or 3 on its item 9 (suicidality). Exclusion criteria were: (1) cannot read or write Japanese; and (2) receiving treatment from mental health professionals at the time of screening.

Study setting. Participants were recruited across Japan from: (1) health insurance societies. There are several types of public health insurance scheme together constituting universal healthcare in Japan. We collaborated with two large associations, the National Federation of Health Insurance Societies covering employees and their families in large corporations (-30 million insured) and the Japan Health Insurance Association covering employees of small- to middle-sized corporations and their families (-35 million insured); (2) business companies and corporations; (3) community and local governments; and (4) direct-to-consumer advertisements.

Procedures, randomization and masking

We asked interested, potential participants to log onto a website specifically prepared for the trial and fill in the screening questionnaire. Once we had ascertained the minimum eligibility criteria (age, possession of smartphone, not receiving treatment from mental health professionals), we invited them to attend the online informed consent session on Zoom in which we explained all the details of the study and its procedures. Once individuals had agreed and signed their digital consent forms, we allowed them access to the app downloaded from the App Store or Google Play by using an ID and password. Consenting individuals then had to complete all the baseline questionnaires within 1 week. After they had completed all the questionnaires, the app automatically randomly allocated participants to 1 of the 12 intervention or control arms according to their employment status (yes versus no) using the preinstalled permuted block randomization sequence, prepared by an independent statistician. The block size was known only to the statistician and allocation was concealed from any study personnel involved in participant recruitment.

Given the nature of the intervention and the study procedure, neither the participants nor the trial management team were blinded to the intervention. All the assessments in this study were completed by the participants on the app.

The participants were remunerated with 1,000 yen for completing the week 3 questionnaires, 2,000 yen at week 6, 1,000 yen at week 26 and 2,000 yen at week 50.

Interventions and controls

Nine intervention arms and one of the three control arms were combined to form the four 2×2 factorial trials, as shown in Fig. 1. The participants were randomized equally to one of the nine intervention arms or three control arms.

Interventions. The intervention arms were: BA, CR, PS, AT, BI, BA + CR, BA + PS, BA + AT and BA + BI. BA consists of psychoeducation about pleasurable activities according to the 'outside-in' principle. It provides a worksheet for a personal experiment to test out a new activity and also a gamified 'action marathon' to promote such personal experiments. CR consists of psychoeducation using the cognitive behavioral model and CR. The participant learns how to monitor their reactions to situations in terms of feelings, thoughts, body reactions and behaviors by filling in mind maps, which the participant then uses to apply CR and find alternative thoughts. PS teaches participants how to break down the issue at hand, to specify a concrete and achievable objective for the issue, to brainstorm possible solutions and compare their advantages and disadvantages, and finally to choose the most desirable action and act on it. A worksheet to guide participants through this process is provided. AT consists of psychoeducation of assertive communication in contrast to aggressive or passive communication. The participant learns how to express their true feelings and wishes without hurting others or sacrificing themselves. They complete worksheets to construct appropriate lines in response to their own real-life interactions. BI teaches the mechanisms of healthy sleep, invites the participant to keep daily sleep records, based on which the participant can start to apply sleep restriction and stimulus control techniques.

Controls. In psychotherapy trials, there is no gold standard control condition such as the pill placebo in pharmacotherapy trials²². Psychotherapy controls must be designed in accordance with the clinical questions of the study as well as the participants' needs and the available resources, and may produce different effect size estimates²³. In the RESiLIENT trial, we therefore used three different control conditions with different levels of stringency: (1) weekly self-checks, (2) health information and (3) delayed treatment.

In the weekly self-check arm, participants received weekly encouragement emails and monitored their moods via weekly PHQ-9 measurements up to week 6 (as in the intervention arms) and then monthly PHQ-9 measurements thereafter up to week 50. This arm is intended as an attention control to match the attention provided through encouragement emails and self-checks but lacking in active ingredients.

In the health information arm, participants received the URLs of websites containing tips for a healthy life (physical activities, nutrition and oral health, none of which focused on mental health) for the initial 3 weeks and then were asked to answer quizzes to test comprehension. Participants were asked to complete self-reports at weeks 3 and 6 (without encouragement emails), and then follow-up evaluations thereafter up to week 50. This arm is intended to emulate a placebo intervention.

In the delayed treatment arm, the participants were placed on the wait list and asked to complete self-reports at weeks 3 and 6. They were then randomized to 1 of the 11 arms (1 to 11) at week 6, if they so wished, and completed the respective programs thereafter.

Resilience Training app. All the interventions and controls were provided on the Resilience Training app v.2.1. Each intervention module consisted of seven chapters, each divided into two to five lessons as appropriate. Extended Data Fig. 1a shows how chapters and lessons were combined in each of the 12 intervention and control arms.

Extended Data Fig. 1b provides screenshots from the smartphone app. Each chapter provided explanations of the skill to be learned as well as a worksheet to practice the learned skill. Participants were expected to complete one chapter per week and could only move on to the next chapter one week after they started the previous chapter and having completed one worksheet.

In addition participants received weekly encouragement emails (except in the health information and delayed treatment arms), which were templated but tailored in accordance with each participant's progress by the trial coordinators in the management team. Participants were prompted to complete the PHQ-9 every week on the app. All the coordinators were forbidden from providing any therapeutic contents.

2×2 factorial trials. In the factorial design, we prioritized BA because it has been shown to be as good as full-package CBT²⁸, was the most effective skill in the component network meta-analysis of iCBT¹⁷ and has increasingly been used in iCBT interventions aiming at scalability as BA alone or in combination of BA plus PS¹⁶. We used delayed treatment as the control arm in the primary analyses but replaced it with the health information and self-check arms in the sensitivity analyses ('Statistical analyses' below).

Concomitant interventions. All participants in the active and control arms were free to seek whatever mental health interventions they wished over the 26 weeks. We recorded the received professional interventions, either pharmacotherapy or psychotherapy at week 6.

Safety monitoring

All serious adverse events were handled according to standard operationalized procedure 10 by the Kyoto University Graduate School of Medicine (http://www.ec.med.kyoto-u.ac.jp/doc/SOP10.pdf). In addition, when any increase in suicidality (defined as a score of 10 or more on the total PHQ-9 and 2 or more on item 9 of PHQ-9 for two consecutive weeks) was noted, an email was sent to the participant recommending that they visit the consultation services listed by the Japanese Ministry of Health, Labour and Welfare for the general public (https://kokoro.mhlw.go.jp/agency/) or other mental health professionals.

Baseline measurements

Demographic, psychosocial and clinical variables were measured at baseline as detailed below. Details of each scale are given in the protocol (Supplementary Information). All measurements were collected on the Resilience Training app.

The measured demographic variables were: (1) age; (2) gender; (3) marital status; (4) number of cohabitants; (5) Area Deprivation Index (as estimated by the zip code in 2020)⁴⁶; (6) education; (7) employment; (8) physical conditions and illnesses, including past and current treatments; (9) physical disability; (10) mental health conditions, including past and current treatments; (11) alcohol use; and (12) the CAGE questionnaire^{47,48}.

The following psychosocial scales were used: (1) the short form of the Big-5 questionnaire⁴⁹⁻⁵¹; (2) Adverse Childhood Experiences^{52,53}; (3) Assessment of Signal Cases (ASC)–Life difficulties⁵⁴; (4) ASC–Social support⁵⁴; (5) CBT Skills Scale⁵⁵; and (6) Readiness for smartphone CBT, (a) ASC–Motivation⁵⁴ and (b) familiarity with smartphone use.

Clinical variables were measured using: (1) PHQ-9⁵⁶, (2) GAD-7⁵⁷, (3) ISI⁵⁸, (4) past and current major depressive episodes according to the major depression section of the computerized WHO Composite International Diagnostic Interview^{59,60}, (5) Work and Social Adjustment Scale⁶¹, (6) Health and Work Performance Questionnaire^{62,63}, (7) Utrecht Work Engagement Scale⁶⁴, (8) EuroQOL-5D-5L^{65,66}, (9) SWEM-WBS^{67,68}, (10) CSQ-3⁶⁹, (11) adherence to smartphone CBT and (12) safety information.

Outcomes

Primary outcome. The primary outcome is the change in PHQ-9 score from baseline to week 6.

Anxiety usually coexists with depression and we aimed to examine what effects the intervention components, although mainly targeting depression, may have on anxiety. We aimed to monitor the broad psychopathology of common mental disorders encompassing depression, anxiety and insomnia as well as positive mental health which deteriorates in common mental disorders.

Sample size

The RESILIENT trial was powered for each factorial trial for its primary outcome, that is, PHQ-9 at week 6. The Fun to Learn, Act and Think through Technology trial showed an SMD of 0.3-0.4 for an earlier version of the Resilience Training app using the combination of BA + CR over the waiting list control among patients with drug-refractory depression⁷¹. To detect an SMD of 0.2 at alpha = 0.05 and beta = 0.10, each factorial trial in the current master protocol would need 1,053 participants in total (FactorialPowerPlan, http://methodology.psu. edu), hence 264 in each arm. Altogether the RESiLIENT trial will require $264 \times 12 = 3,168$ participants with baseline PHQ-9 scores of 5 or more. The Healthy Campus Trial, which used an earlier version of the Resilience Training app among university students, had 5% dropout rates for its eighth week outcome²⁵. We expected somewhat higher dropout rates in the RESiLIENT trial, which targets the general adult population in the community. Assuming a 10% dropout rate, the required sample size was 3.520.

Statistical analyses

We conducted all the analyses with R (v.4.3.1), and used the packages mmrm (v.0.3.11)⁷² and emmeans (v.1.9.0)⁷³.

Efficacy analyses of the primary outcome. *Primary efficacy analyses.* In the preplanned primary analysis, each of the four 2 × 2 factorial trials was analyzed independently in accordance with the master protocol design framework. The four trials used the BA arm (intervention arm 1) and the control arm (intervention arm 12, delayed treatment) in common. Intervention arms 11 (health information) and 10 (self-check), with increasing stringency as the control, were used in sensitivity analyses; that is, we performed additional analyses changing the control arm to arm 11 or 10 respectively. For the fourth trial involving BA and BI, participants who worked in three shifts were excluded because shift workers cannot be expected to make optimal use of BI.

We used the MMRM^{74,75} to estimate the mean difference in change scores on the PHQ-9 for each component while adjusting for missing outcomes. The model included fixed effects of treatment, visit (as categorical) and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. Each of the experimental factors was coded at two levels (presence was coded as 1 and absence was coded as 0). The covariance matrix structure of outcome variables was set to unstructured. The standard error estimates and degrees of freedom were adjusted using the Kenward-Roger method^{76,77}. We also estimated least squares mean change scores for the PHQ-9 at weeks 1, 2, 3, 4 and 5 based on the estimated models of the MMRM analysis. For the effect of BA, which was estimated in all four 2 × 2 trials, we presented the effect size estimate of the first trial (the 2 × 2 trial involving BA and CR) as the primary result because BA and CR are the most typical ingredients across broadly conceived CBT interventions.

We applied no adjustment for multiple testing in examination of statistical significance of the main effects and the conventional threshold for statistical significance (P < 0.05, two-sided), because the multiple hypotheses have been tested independently in these trial designs

assuming as if the tested interventions could have been assessed in separate trials^{78,79}. We performed another MMRM analysis involving a two-way interaction term of each two components to assess the interaction between the treatments as a sensitivity analysis. Previous research has not been suggestive of interactions among iCBT components^{17,25}, but when we identified a strong interaction, we interpreted the results considering the interaction.

Effect sizes (SMDs) were calculated using the observed baseline SD for within-group change scores and the observed week 6 SD for group differences.

Secondary efficacy analyses. To assess the comparative efficacy of the five CBT skills and their combinations, we additionally performed a combined analysis via pooling the four factorial trials along with all the control arms. For this we performed the MMRM analysis for the pooled dataset and provided comparative effect size information for all the CBT skills and their combinations evaluated in the four trials. In this analysis we included all the three control conditions and compared comprehensively. In this combined analysis, we included shift workers, although such inclusion may lead to underestimating the effect of BI on depression, to maintain the statistical power for the analysis and make the results as comparable as possible to those in the primary 2×2 trial analyses.

Efficacy analyses for secondary outcomes. We performed MMRM analyses for the repeated measurement outcomes of GAD-7, ISI and SWEMWBS until week 6. We used the same modeling strategy with the analysis of the primary outcome.

Sensitivity analyses. We evaluated the robustness of the foregoing analyses by conducting the following prespecified sensitivity analyses. For the 2×2 trials:

- (1) We examined interactions among the skills as a sensitivity analysis. In examining interactions among components, we also performed linear regression analyses using only the week 6 data for each of the 2 × 2 trials.
- (2) We used control arm 10 (self-check) or 11 (health information) instead of 12 (delayed treatment) in each of the 2 × 2 trials.
- (3) For the factorial trial examining BA and BI, we subgrouped participants into those who scored ≥8 versus those who scored ≤7 on the ISI at baseline. Here, the primary outcome was the PHQ-9, and this subgroup analysis confirmed whether the effect of BI is similar regardless of the baseline insomnia severity. For the secondary outcomes:
- (4) For the secondary outcome of the GAD-7, we limited participants to those who scored ≥5 on the GAD-7 at baseline (that is, those who presented with at least some anxiety symptoms at baseline). This mitigated the floor effect on the GAD-7.
- (5) For the secondary outcome of the ISI, we limited participants to those who scored ≥8 on the ISI at baseline (that is, those who presented with at least some insomnia symptoms at baseline). This mitigated the floor effect on the ISI.
- (6) For the secondary outcome of the SWEMWBS in the four 2×2 factorial trials, we limited participants to those who scored ≤20 on the SWEMWBS at baseline (that is, those who showed reduced wellbeing at baseline). This mitigated the ceiling effect on the SWEMWBS.

Adherence. We used two definitions of adherence to the program in accordance with the previous literature: completion of the whole program and completion of the main parts of the program^{29,30}. In this study, we operationally defined the former as completing all seven chapters (the introduction, chapters 1 to 5 and the epilogue) and the required worksheets, and the latter as finishing at least up to chapter

4 and the associated worksheets because the Introduction and the first four chapters cover the main learning materials to be taught in each arm, whereas chapter 5 and the epilogue are only review materials (Extended Data Fig. 1a). We set the deadline for the acute phase intervention at week 10 and that for the follow-up at week 30 to allow participants some additional time beyond the program schedule to complete any missed lessons or tasks.

Safety analyses. We descriptively reported potential adverse reactions and serious adverse events during the 6-week intervention via user self-reports.

Lived experience involvement

We revised the texts and the user interface of the app in collaboration with a group of people with lived experiences. However, they were not involved in the trial design, data analyses or interpretations of the current study findings.

Changes from the protocol or the SAP

The published protocol (Supporting Information) declared that we had two analyses planned for the RESiLIENT trial, focusing on the change in PHQ-9 from baseline to week 6 (analysis 1) and the incidence of a major depressive episode by week 50 (analysis 2). This paper is a report for analysis 1, and we prespecified the SAP accordingly before the relevant data became available (Supplementary Information). We will report analysis 2 and the outcomes listed as secondary in the published protocol, but not covered in the current SAP, in future papers. We plan to publish a SAP for the latter analyses before the data for the one-year outcomes become available.

In this study, we followed the analyses as prespecified in the SAP for analysis 1 (Supplementary Information) but added two post hoc analyses to inform and facilitate interpretations of the primary analysis results.

First, we extended the MMRM analyses of the primary outcome (PHQ-9) up to week 26 to test the durability of the effects we had confirmed immediately after the interventions at week 6. The statistical power for these additional analyses remained adequate for an SMD of 0.2 at alpha = 0.05 and beta = 0.10 assuming a 10% dropout rate, as originally required for the week 6 primary analyses of the 2×2 factorial trials.

Second, we applied the analytical methods of component analyses to the week 6 and week 26 outcomes to more finely differentiate the specific efficacies of individual CBT skills in addition to the nonspecific treatment effects and the waiting list effect based on the MMRM model. In this model, we considered the health information control, in which participants read information on physical, nutritional and oral health on the app, but received no encouragement emails and did not fill in the PHQ-9 on a weekly basis, as the reference. In comparison with this, the delayed treatment control was considered to have the waiting list component $wl^{17,23,35}$. All the intervention arms were assumed to include a nonspecific treatment component *ns*, which included the placebo effect, human contact via encouragement emails and weekly self-checks^{17,35}, in addition to their specific skill effects.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Deidentified individual participant data and data dictionary will be made available 24 months after publication on UMIN-ICDR, an individual case data repository managed by the Japanese University Hospital Medical Information Network (UMIN) Center (https://www.umin.ac.jp/ icdr/index.html). Proposals with specific aims and an analysis plan should be directed to the corresponding author, T.A.F. (furukawa@ kuhp.kyoto-u.ac.jp). The decision will be made within a month.

Code availability

R code files used in the statistical analyses are available on GitHub at https://github.com/nomahi/RESiLIENT.

References

- Schulz, K. F., Altman, D. G. & Moher, D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c332 (2010).
- Boutron, I. et al. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann. Intern. Med.* 167, 40–47 (2017).
- Kahan, B. C. et al. Reporting of factorial randomized trials: extension of the CONSORT 2010 statement. JAMA **330**, 2106–2114 (2023).
- Spitzer, R. L., Kroenke, K. & Williams, J. B. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 282, 1737–1744 (1999).
- Nakaya, T. et al. Associations of all-cause mortality with censusbased neighbourhood deprivation and population density in Japan: a multilevel survival analysis. *PLoS ONE* 9, e97802 (2014).
- Fiellin, D. A., Reid, M. C. & O'Connor, P. G. Screening for alcohol problems in primary care: a systematic review. *Arch. Intern. Med.* 160, 1977–1989 (2000).
- Bush, B., Shaw, S., Cleary, P., Delbanco, T. L. & Aronson, M. D. Screening for alcohol abuse using the CAGE questionnaire. *Am. J. Med.* 82, 231–235 (1987).
- Wada, S. Construction of the Big Five Scales of personality trait terms and concurrent validity with NPI [in Japanese]. Shinrigaku Kenkyu 67, 61–67 (1996).
- 50. Namikawa, T. et al. Development of a short form of the Japanese Big-Five Scale, and a test of its reliability and validity [in Japanese]. Shinrigaku Kenkyu **83**, 91–99 (2012).
- 51. Toyomoto, R. et al. Validation of the Japanese Big Five Scale Short Form in a university student sample. *Front. Psychol.* **13**, 862646 (2022).
- 52. Felitti, V. J. et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *Am. J. Prev. Med.* **14**, 245–258 (1998).
- Tsuboi, S. Creating a Japanese version Adverse Childhood Experiences score calculator [in Japanese]. in KAKENHI Reports (2014); https://kaken.nii.ac.jp/ja/report/KAKENHI-PROJECT-24790625/24790625seika/
- 54. Probst, T., Lambert, M. J., Dahlbender, R. W., Loew, T. H. & Tritt, K. Providing patient progress feedback and clinical support tools to therapists: is the therapeutic process of patients on-track to recovery enhanced in psychosomatic in-patient therapy under the conditions of routine practice? J. Psychosom. Res **76**, 477–484 (2014).
- 55. Sakata, M. et al. Development and validation of the cognitive behavioural therapy skills scale among college students. *Evid. Based Ment. Health* **24**, 70–76 (2021).
- Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med* 16, 606–613 (2001).
- Spitzer, R. L., Kroenke, K., Williams, J. B. & Lowe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166, 1092–1097 (2006).
- Bastien, C. H., Vallières, A. & Morin, C. M. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2, 297–307 (2001).

- Peters, L. & Andrews, G. Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. *Psychol. Med.* 25, 1269–1280 (1995).
- Shimoda, H., Inoue, A., Tsuno, K. & Kawakami, N. One-year testretest reliability of a Japanese web-based version of the WHO Composite International Diagnostic Interview (CIDI) for major depression in a working population. *Int. J. Methods Psychiatr. Res.* 24, 204–212 (2015).
- Mundt, J. C., Marks, I. M., Shear, M. K. & Greist, J. H. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br. J. Psychiatry* **180**, 461–464 (2002).
- Kessler, R. C. et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). J. Occup. Environ. Med. 45, 156–174 (2003).
- Kawakami, N. et al. Construct validity and test-retest reliability of the World Mental Health Japan version of the World Health Organization Health and Work Performance Questionnaire Short Version: a preliminary study. *Ind. Health* 58, 375–387 (2020).
- 64. Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M. & De Witte, H. An ultra-short measure for work engagement. *Eur. J. Psychol. Assess.* **35**, 577–591 (2017).
- 65. Herdman, M. et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* **20**, 1727–1736 (2011).
- Ikeda, S. et al. Developing a Japanese version of the EQ-5D-5L value set. J. Natl Inst. Public Health 64, 47–55 (2015).
- 67. Bartram, D. J., Sinclair, J. M. & Baldwin, D. S. Further validation of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis. *Qual. Life Res.* **22**, 379–391 (2013).
- Shah, N., Cader, M., Andrews, B., McCabe, R. & Stewart-Brown, S. L. Short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health Qual. Life Outcomes* 19, 260 (2021).
- Attkisson, C. C. & Zwick, R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval. Program Plann.* 5, 233–237 (1982).
- Morin, C. M., Belleville, G., Bélanger, L. & Ivers, H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* 34, 601–608 (2011).
- 71. Mantani, A. et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. *J. Med. Internet Res.* **19**, e373 (2017).
- 72. Sabanes Bove, D. et al. mmrm: Mixed Models for Repeated Measures (2024); https://doi.org/10.32614/CRAN.package.mmrm
- 73. Lenth, R. emmeans: Estimated Marginal Means, aka Least-Squares Means (2023); https://doi.org/10.32614/CRAN.package.emmeans
- Mallinckrodt, C. H., Clark, W. S. & David, S. R. Accounting for dropout bias using mixed-effects models. J. Biopharm. Stat. 11, 9–21 (2001).
- 75. Mallinckrodt, C. H., Kaiser, C. J., Watkin, J. G., Molenberghs, G. & Carroll, R. J. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin. Trials* 1, 477–489 (2004).
- Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53, 983–997 (1997).
- 77. Kenward, M. G. & Roger, J. H. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Comput. Stat. Data Anal.* **53**, 2583–2595 (2009).

- Balakrishnan, N. Methods and Applications of Statistics in Clinical Trials, Volume 1 and Volume 2: Concepts, Principles, Trials, and Designs (Wiley, 2014).
- 79. Fleiss, J. L. Design and Analysis of Clinical Experiments (Wiley, 1999).

Acknowledgements

This study was supported by the Japan Agency for Medical Research and Development (AMED) (grant no. JP21de0107005), awarded to T.A.F. (principal investigator), A.T., R.T., M.S., Y.L., M.H., T.A., N. Kawakami, T.N., N. Kondo and S.F. The funder had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Author contributions

T.A.F. and M.H. conceived the study. T.A.F., J.M.S.W. and H.N. designed the study. Additional advice about design and statistical analysis was given by R.C.K., H.C. and P.C. Funding was obtained by T.A.F., M.H., T.A., N. Kawakami, T.N., N. Kondo, S.F. and H.N. T.A.F., M.H. and T.A. led the software development, with input from A.T., R.T., M.S. and Y.L. A.T., R.T., M.S. and Y.L. established the databases and managed the recruitment, data collection and processing, with help from N. Kawakami, T.N., N. Kondo and S.F. H.N. and T.A.F. undertook the statistical analyses, and had direct access to the full data. T.A.F. and H.N. wrote the first draft, and all the authors contributed to the interpretation and critical revision of the manuscript, and approved the final paper.

Competing interests

T.A.F. reports personal fees from Boehringer Ingelheim, Daiichi Sankyo, DT Axis, Micron, Shionogi, SONY and UpToDate, and a grant from DT Axis and Shionogi, outside the submitted work. In addition, T.A.F. has a patent (7448125) and a pending patent (2022-082495), and has licensed intellectual properties for Kokoro-app to DT Axis. A.T. reports personal fees from Eisai and Shionogi outside the submitted work. M.S. is employed in the Department of Neurodevelopmental Disorders, Nagoya City University Graduate School of Medical Sciences, which is an endowment department supported by the City of Nagoya, and has received a personal fee from SONY outside the submitted work. M.H. has a patent (7448125) and has licensed intellectual properties for Kokoro-app to Mitsubishi Tanabe. T.A. has received lecture fees from AstraZeneca, Chugai, Dajichi Sankvo, Eisai, Eli Lilly, Kowa, Kyowa Kirin, Lundbeck, MSD, Meiji-Seika Pharma, Merck, Nipro, Otsuka, Pfizer, Shionogi, Sumitomo pharma, Takeda, Tsumura, UCB and Viatris, and a grant from Shinogi and royalties from Igaku-Shoin, outside the submitted work. T.A. serves as the representative director of the General Incorporated Association

NCU CRESS and receives compensation as an advisor to Snom Inc. T.A. has pending patents (2020-135195 and 2024-516288) (Institute) and patents (7313617) (Institute). N. Kawakami is employed by the Junpukai Foundation and the Department of Digital Mental Health, The University of Tokyo, an endowment department that is supported by an unrestricted grant from 15 enterprises (https://dmh.m.u-tokyo. ac.jp/en). T.N. reports grants or contracts with I&H, Cocokarafine, Konica Minolta and NTT DATA, consulting fees from Ohtsuka, Takeda, Johnson & Johnson, AstraZeneca and Nippon Zoki, payment or honoraria for lectures, manuscript writing or educational events from Pfizer, MSD, Chugai, Takeda, Janssen, Boehringer Ingelheim, Eli Lilly, Maruho, Mitsubishi Tanabe, Novartis, Allergan, Novo Nordisk, Toa Eiyo, AbbVie, Ono, GSK, Alexion, Cannon Medical Systems, Kowa, Araya, Merck and Amicus, stock options in BonBon Inc., and a donation from Cancerscan and JMDC, all outside the submitted work. S.F. received personal fees from Kyowa Kirin, Boehringer Ingelheim, Toray Medical, Health and Global Policy Institute, White Healthcare and Health Insurance for Start-up Companies, and research grants from the Japan Health Insurance Association, National Health Insurance Association for Civil Engineering and Construction, Osaka Prefecture and Cancerscan, outside this work. R.C.K. was a consultant for Cambridge Health Alliance, Canandaigua VA Medical Center, Child Mind Institute, Holmusk, Massachusetts General Hospital, Partners Healthcare, Inc., RallyPoint Networks, Inc., Sage Therapeutics and University of North Carolina. He has stock options in Cerebral Inc., Mirah, PYM (Prepare Your Mind), Roga Sciences and Verisense Health. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-025-03639-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-025-03639-1.

Correspondence and requests for materials should be addressed to Toshi A. Furukawa.

Peer review information *Nature Medicine* thanks Thole Hoppen, Zhirong Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ming Yang, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

		Introduction	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Epilogue
1	BA	3 lessons for	4 basic	2 advanced	1 advanced	1 advanced	BA in everyday	Epilogue
		psychoeducation	lessons for BA	lessons for BA	lesson for BA	lesson for BA	life	
2	CR	3 lessons for	4 basic	2 advanced	1 advanced	1 advanced	CR in everyday	Epilogue
		psychoeducation	lessons for CR	lessons for CR	lesson for CR	lesson for CR	life	
3	PS	3 lessons for	4 basic	2 advanced	2 advanced	1 advanced	PS in everyday	Epilogue
		psychoeducation	lessons for PS	lessons for PS	lessons for PS	lesson for PS	life	
4	AT	3 lessons for	5 basic	2 advanced	1 advanced	1 advanced	AT in everyday	Epilogue
		psychoeducation	lessons for AT	lessons for AT	lesson for AT	lesson for AT	life	
5	BI	3 lessons for	5 basic	1 advanced	1 advanced	1 advanced	BI in everyday	Epilogue
		psychoeducation	lessons for BI	lesson for Bl	lesson for B	lesson for Bl	life	
6	BA+CR	3 lessons for	4 basic	4 advanced	4 basic	4 advanced	BA&CR in	Epilogue
		psychoeducation	lessons for BA	lessons for BA	lessons for CR	lessons for CR	everyday life	
7	BA+PS	3 lessons for	4 basic	4 advanced	4 basic	5 advanced	BA&PS in	Epilogue
		psychoeducation	lessons for BA	lessons for BA	lessons for PS	lessons for PS	everyday life	
8	BA+AT	3 lessons for	4 basic	4 advanced	5 basic	4 advanced	BA&AT in	Epilogue
		psychoeducation	lessons for BA	lessons for BA	lessons for AT	lessons for AT	everyday life	
9	BA+BI	3 lessons for	4 basic	4 advanced	5 basic	3 advanced	BA&BI in	Epilogue
		psychoeducation	lessons for BA	lessons for BA	lessons for BI	lessons for BI	everyday life	
10	Self-check	1 lesson for	-	-	-	-	-	Epilogue
		psychoeducation						
11	Health	1 lesson for	Physical	Nutrition	Oral health	-	-	Epilogue
	information	psychoeducation	exercise					
12	Delayed	1 lesson for	-	-	-	-	-	Epilogue
	treatment	psychoeducation						

1a. Proceedings of the intervention and control arms

1b. Screenshots of the five skills of the Resilience Training App Version 2.1.



$\label{eq:contents} Extended \, Data \, Fig. \, 1 | \, Contents \, of the interventions \, and \, controls.$

The intervention arms #1 through #10 (BA through Self-check) filled in the PHQ-9 every week and received weekly encouragement emails from the clinical research coordinators at the central office. #11 (Health information) and #12

(Delayed treatment) were asked to fill in the PHQ-9 at weeks 3 and 6 only, and did not receive encouragement emails. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia CR: Cognitive restructuring, PS: Problem solving.



Extended Data Fig. 2 | PHQ-9 change from baseline up to week 26. Data represent least squares mean changes of the PHQ-9 scores estimated by the combined MMRM model. The regression model included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores employment status, age and gender. The error bars

at weeks 6 and 26 represent SEMs. For more information, see Table 4, Extended Data Table 8, and Supplementary Table 2. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving.

BA

CR PS AT BI

Extended Data Table 1 | Analysis of main and interaction effects using the mixed-model repeated measures analysis (MMRM) for PHQ-9 score at week 6, using the Health information arm as the control

			Differen	ces in change scores of th	ne PHQ-9 at W	'eek 6
Trial *	A #177	NI	Estimated least	Effect size [‡]	P-value	P-value
mat "	AIIII	IN	squares mean		(Main	(Interaction of
			differences [†]		effect)	two components)
Trial 1	BA + CR	329	-1.58 (-2.06 to -1.10)	-0.47 (-0.61 to -0.33)	1.3×10 ⁻¹⁰	1.9×10 ⁻⁴
	BA	327	-1.63 (-2.11 to -1.15)	-0.49 (-0.63 to -0.34)	3.9×10 ⁻¹¹	
	CR	328	-1.25 (-1.72 to -0.77)	-0.37 (-0.51 to -0.23)	3.4×10 ⁻⁷	
	Health information [reference]	328	—	—	—	
Trial 2	BA + PS	329	-1.69 (-2.17 to -1.22)	-0.51 (-0.65 to -0.36)	3.3×10 ⁻¹²	0.001
	BA	327	-1.64 (-2.11 to -1.16)	-0.49 (-0.63 to -0.35)	2.2×10 ⁻¹¹	
	PS	328	-1.18 (-1.66 to -0.71)	-0.35 (-0.49 to -0.21)	1.2×10 ⁻⁶	
	Health information [reference]	328	—	—	—	
Trial 3	BA + AT	328	-1.50 (-2.00 to -1.01)	-0.43 (-0.57 to -0.29)	3.8×10 ⁻⁹	1.1×10 ⁻⁴
	BA	327	-1.63 (-2.13 to -1.13)	-0.47 (-0.61 to -0.32)	1.7×10 ⁻¹⁰	
	AT	328	-1.26 (-1.75 to -0.76)	-0.36 (-0.50 to -0.22)	7.6×10 ⁻⁷	
	Health information [reference]	328	—	—	—	
Trial 4**	BA + BI	311	-1.39 (-1.88 to -0.90)	-0.40 (-0.54 to -0.26)	2.7×10 ⁻⁸	3.1×10 ⁻⁷
	BA	314	-1.63 (-2.12 to -1.14)	-0.47 (-0.61 to -0.33)	8.1×10 ⁻¹¹	
	BI	313	-1.56 (-2.05 to -1.08)	-0.45 (-0.59 to -0.31)	3.7×10 ⁻¹⁰	
	Health information [reference]	313	—	—	—	

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression models included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. The interaction P-values are calculated from the MMRM models involving interaction terms of the corresponding treatments. P-values are two-sided and no adjustments are made for multiple comparisons. † Minus value indicates depression reduction. () indicates 95% confidence intervals. ‡ Effect sizes (standardised mean differences) were calculated using the observed week 6SD for group differences. * See Fig. 1 for the designs of Trials 1 through 4. ** Participants working on a three-shift schedule were excluded in the primary and its sensitivity analyses for Trial 4. AT: Assertion training, BA: Behavioural activation, BI: Behaviour Herapy for insomnia, CR: Cognitive restructuring, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving.

Extended Data Table 2 | Analysis of main and interaction effects using the mixed-model repeated measures analysis (MMRM) for PHQ-9 score at week 6, using the Self-check arm as the control

			Differences in change scores of the PHQ-9 at Week 6							
Trial *	٨٢٣	N	Estimated least	Effect size [‡]	P-value	<i>P</i> -value				
mat "	AIIII	IN	squares mean		(Main	(Interaction of				
			differences [†]		effect)	two components)				
Trial 1	BA + CR	329	-0.96 (-1.44 to -0.47)	-0.28 (-0.43 to -0.14)	1.1×10 ⁻⁴	0.068				
	BA	327	-0.99 (-1.47 to -0.50)	-0.29 (-0.43 to -0.15)	7.1×10 ⁻⁵					
	CR	328	-0.61 (-1.09 to -0.13)	-0.18 (-0.32 to -0.04)	0.014					
	Self-check [reference]	328	—	—	—					
Trial 2	BA + PS	329	-1.06 (-1.54 to -0.58)	-0.31 (-0.45 to -0.17)	1.5×10 ⁻⁵	0.163				
	BA	327	-1.00 (-1.48 to -0.52)	-0.29 (-0.44 to -0.15)	5.2×10 ⁻⁵					
	PS	328	-0.55 (-1.03 to -0.07)	-0.16 (-0.30 to -0.02)	0.026					
	Self-check [reference]	328	—	—	—					
Trial 3	BA + AT	328	-0.88 (-1.38 to -0.37)	-0.25 (-0.39 to -0.11)	0.001	0.039				
	BA	327	-1.00 (-1.50 to -0.49)	-0.28 (-0.42 to -0.14)	1.0×10 ⁻⁴					
	AT	328	-0.63 (-1.13 to -0.13)	-0.18 (-0.32 to -0.04)	0.014					
	Self-check [reference]	328	—	—	—					
Trial 4**	BA + BI	311	-0.77 (-1.26 to -0.28)	-0.22 (-0.36 to -0.08)	0.002	9.0×10 ⁻⁴				
	BA	314	-1.01 (-1.50 to -0.51)	-0.29 (-0.43 to -0.15)	6.8×10 ⁻⁵					
	BI	313	-0.94 (-1.44 to -0.45)	-0.27 (-0.41 to -0.13)	1.7×10 ⁻⁴					
	Self-check [reference]	313	—	—	—					

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression models included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. The interaction P-values are calculated from the MMRM models involving interaction terms of the corresponding treatments. P-values are two-sided and no adjustments are made for multiple comparisons. † Minus value indicates depression reduction. () indicates 95% confidence intervals. ‡ Effect sizes (standardised mean differences) were calculated using the observed week 6SD for group differences. * See Fig. 1 for the designs of Trials 1 through 4. ** Participants working on a three-shift schedule were excluded in the primary and its sensitivity analyses for Trial 4. AT: Assertion training, BA: Behavioural activation, BI: Behaviour Herapy for insomnia, CR: Cognitive restructuring, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving.

Extended Data Table 3 | Combined analysis using the mixed-model repeated measures analysis (MMRM) for GAD-7 score at week 6

Arm	N	Estimated least squares mean change scores of the PHQ-9 [†]					HQ-9†	Effect size for Baseline-Week 6 change scores [‡]	Estimated differences of change scores (vs DT)		Effect size for differences of change scores [†] (95% CI)			
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	(95% CI)	(95% CI)	P-value	vs DT	vs HI	vs SC	
ВА	327	-	-	-1.71	-	-	-2.42	-0.71 (-0.86 to -0.56)	-1.66 (-2.10 to -1.21)	5.1×10 ⁻¹³	-0.49 (-0.62 to -0.36)	-0.24 (-0.37 to -0.10)	-0.38 (-0.52 to -0.25)	
CR	328	_	_	- 1.53	—	_	- 1.84	-0.54 (-0.69 to -0.39)	-1.08 (-1.53 to -0.64)	1.9×10 ⁻⁶	-0.32 (-0.45 to -0.19)	-0.07 (-0.20 to 0.06)	-0.21 (-0.35 to -0.08)	
PS	328	_	-	-1.16	_	_	-1.63	-0.48 (-0.62 to -0.33)	-0.87 (-1.31 to -0.42)	1.5×10 ⁻⁴	-0.26 (-0.39 to -0.12)	0.00 (-0.14 to 0.13)	-0.15 (-0.28 to -0.02)	
AT	328	_	-	-1.23	_	_	- 1.75	-0.51 (-0.66 to -0.37)	-0.99 (-1.44 to -0.55)	1.3×10 ⁻⁵	-0.29 (-0.43 to -0.16)	-0.04 (-0.17 to 0.09)	-0.19 (-0.32 to -0.06)	
ВІ	328	-	-	-1.89	-	-	-2.33	-0.68 (-0.83 to -0.54)	-1.57 (-2.02 to -1.13)	5.4×10 ⁻¹²	-0.47 (-0.60 to -0.33)	-0.21 (-0.34 to -0.08)	-0.36 (-0.49 to -0.23)	
BA + CR	329	-	-	-1.80	-	-	- 2.10	-0.62 (-0.76 to -0.47)	-1.34 (-1.79 to -0.89)	4.7×10 ⁻⁹	-0.40 (-0.53 to -0.26)	-0.14 (-0.28 to -0.01)	-0.29 (-0.42 to -0.16)	
BA + PS	329	-	-	-1.76	_	-	- 2.16	-0.63 (-0.78 to -0.49)	-1.40 (-1.84 to -0.95)	8.5×10 ⁻¹⁰	-0.41 (-0.55 to -0.28)	-0.16 (-0.29 to -0.03)	-0.31 (-0.44 to -0.18)	
BA + AT	328	-	-	- 1.61	-	-	- 2.03	-0.59 (-0.74 to -0.45)	-1.27 (-1.71 to -0.82)	2.8×10 ⁻⁸	-0.38 (-0.51 to -0.24)	-0.12 (-0.26 to 0.01)	-0.27 (-0.40 to -0.14)	
BA + BI	327	-	-	- 1.50	-	-	- 2.02	-0.59 (-0.74 to -0.44)	-1.26 (-1.70 to -0.81)	3.7×10 ⁻⁸	-0.37 (-0.50 to -0.24)	-0.12 (-0.25 to 0.01)	-0.27 (-0.40 to -0.13)	
Self-check [reference]	328	_	_	-1.02	_	_	-1.12	-0.33 (-0.47 to -0.18)	-0.36 (-0.80 to 0.09)	0.114	-0.11 (-0.24 to 0.03)	0.15 (0.02 to 0.28)	_	
Health information [reference]	328	_	_	-1.43	_	_	-1.62	-0.47 (-0.62 to -0.33)	-0.85 (-1.30 to -0.41)	1.7×10 ⁻⁴	-0.25 (-0.38 to -0.12)	_	-0.15 (-0.28 to -0.02)	
Delayed treatment [reference]	328	-	-	-0.79	-	-	-0.76	-0.22 (-0.37 to -0.08)	_	_	_	0.25 (0.12 to 0.38)	0.11 (-0.03 to 0.24)	

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression model included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. P-values are two-sided and no adjustments are made for multiple comparisons. † Minus value indicates depression reduction. () indicates 95% confidence intervals. ‡ Effect sizes (standardised mean differences) were calculated using the observed baseline SD for within-group change scores and the observed week 6 SD for group differences. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, DT: Delayed treatment, HI: Health information, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving, SC: Self-check.

Extended Data Table 4 | Combined analysis using the mixed-model repeated measures analysis (MMRM) for ISI score at week 6

Arm	N	Estimated least squares mean change scores of the PHQ-9 [†]					PHQ-9 [†]	Effect size for Baseline-Week 6 change scores [‡]	Estimated diffe of change so (vs DT)	erences cores	Effect size for differences of change scores [†] (95% CI)			
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	(95% CI)	(95% CI)	P-value	vs DT	vs HI	vs SC	
ВА	327	-	-	-0.90	_	_	-1.63	-0.36 (-0.50 to -0.23)	-1.18 (-1.75 to -0.61)	5.3×10 ⁻⁵	-0.25 (-0.37 to -0.13)	-0.09 (-0.21 to 0.03)	-0.22 (-0.34 to -0.10)	
CR	328	-	-	-0.90	-	-	-1.36	-0.30 (-0.44 to -0.16)	-0.91 (-1.48 to -0.34)	0.002	-0.19 (-0.31 to -0.07)	-0.03 (-0.15 to 0.09)	-0.16 (-0.28 to -0.04)	
PS	328	-	-	-0.79	-	-	-1.14	-0.25 (-0.39 to -0.12)	-0.69 (-1.27 to -0.12)	0.018	-0.15 (-0.27 to -0.03)	0.01 (-0.11 to 0.13)	-0.12 (-0.24 to 0.00)	
AT	328	_	-	-0.87	_	-	-1.42	-0.32 (-0.45 to -0.18)	-0.97 (-1.54 to -0.40)	0.001	-0.21 (-0.33 to -0.08)	-0.05 (-0.17 to 0.07)	-0.18 (-0.30 to -0.06)	
ВІ	328	_	-	-2.10	_	_	- 2.77	-0.62 (-0.75 to -0.48)	-2.32 (-2.89 to -1.75)	1.7×10 ⁻¹⁵	-0.49 (-0.61 to -0.37)	-0.33 (-0.45 to -0.21)	-0.46 (-0.58 to -0.34)	
BA + CR	329	-	-	-1.08	-	-	-1.60	-0.36 (-0.50 to -0.22)	-1.15 (-1.72 to -0.58)	7.8×10 ⁻⁵	-0.24 (-0.36 to -0.12)	-0.09 (-0.21 to 0.04)	-0.22 (-0.34 to -0.09)	
BA + PS	329	_	_	-1.17	_	_	-1.61	-0.36 (-0.50 to -0.22)	-1.16 (-1.73 to -0.60)	6.0×10 ⁻⁵	-0.25 (-0.37 to -0.13)	-0.09 (-0.21 to 0.03)	-0.22 (-0.34 to -0.10)	
BA + AT	328	_	-	-1.13	_	_	-1.63	-0.36 (-0.50 to -0.22)	-1.18 (-1.75 to -0.61)	5.2×10 ⁻⁵	-0.25 (-0.37 to -0.13)	-0.09 (-0.21 to 0.03)	-0.22 (-0.34 to -0.10)	
BA + BI	327	_	-	-1.32	_	_	-2.46	-0.55 (-0.68 to -0.41)	-2.01 (-2.58 to -1.44)	5.5×10 ⁻¹²	-0.43 (-0.55 to -0.30)	-0.27 (-0.39 to -0.15)	-0.40 (-0.52 to -0.28)	
Self-check [reference]	328	_	-	-0.35	_	_	-0.59	-0.13 (-0.27 to 0.00)	-0.14 (-0.70 to 0.43)	0.639	-0.03 (-0.15 to 0.09)	0.13 (0.01 to 0.25)	_	
Health information [reference]	328	_	-	-0.83	_	_	-1.20	-0.27 (-0.40 to -0.13)	-0.75 (-1.31 to -0.18)	0.010	-0.16 (-0.28 to -0.04)	_	-0.13 (-0.25 to -0.01)	
Delayed treatment [reference]	328	-	-	-0.40	-	-	-0.45	-0.10 (-0.24 to 0.04)	_	_	_	0.16 (0.04 to 0.28)	0.03 (-0.09 to 0.15)	

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression model included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. P-values are two-sided and no adjustments are made for multiple comparisons. † Minus value indicates depression reduction. () indicates 95% confidence intervals. ‡ Effect sizes (standardised mean differences) were calculated using the observed baseline SD for within-group change scores and the observed week 6 SD for group differences. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, DT: Delayed treatment, HI: Health information, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving, SC: Self-check.

Extended Data Table 5 | Combined analysis using the mixed-model repeated measures analysis (MMRM) for SWEMWBS score at week 6

Arm	N	Estimated least squares mean change scores of the PHQ-91					'HQ-9 [†]	Effect size for Baseline-Week 6 change scores [‡]	Estimated diff of change s (vs DT)	erences cores	Effect size for differences of change scores [†] (95% CI)			
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	(95% CI)	(95% CI)	P-value	vs DT	vs HI	vs SC	
ВА	327	-	_	0.42	—	_	0.97	0.26 (0.11 to 0.41)	1.11 (0.60 to 1.63)	2.0×10 ⁻⁵	0.26 (0.14 to 0.38)	0.09 (-0.03 to 0.21)	0.15 (0.03 to 0.27)	
CR	328	_	_	0.19	-	_	0.62	0.17 (0.02 to 0.32)	0.77 (0.26 to 1.28)	0.003	0.18 (0.06 to 0.30)	0.01 (-0.11 to 0.13)	0.07 (-0.04 to 0.19)	
PS	328	-	-	-0.15	-	-	0.18	0.05 (-0.10 to 0.20)	0.32 (-0.19 to 0.84)	0.214	0.08 (-0.04 to 0.19)	-0.09 (-0.21 to 0.02)	-0.03 (-0.15 to 0.09)	
AT	328	_	-	0.15	_	_	0.57	0.16 (0.00 to 0.31)	0.72 (0.21 to 1.23)	0.006	0.17 (0.05 to 0.28)	0.00 (-0.12 to 0.11)	0.06 (-0.06 to 0.18)	
ВІ	328	_	-	0.57	_	-	1.03	0.28 (0.13 to 0.43)	1.18 (0.67 to 1.69)	5.8×10 ⁻⁶	0.27 (0.15 to 0.39)	0.10 (-0.02 to 0.22)	0.17 (0.05 to 0.29)	
BA + CR	329	-	-	0.15	_	-	0.53	0.14 (-0.01 to 0.30)	0.67 (0.16 to 1.19)	0.010	0.16 (0.04 to 0.27)	-0.01 (-0.13 to 0.10)	0.05 (-0.07 to 0.17)	
BA + PS	329	_	_	0.32	_	-	0.63	0.17 (0.02 to 0.32)	0.77 (0.27 to 1.28)	0.003	0.18 (0.06 to 0.30)	0.01 (-0.11 to 0.13)	0.07 (-0.04 to 0.19)	
BA + AT	328	_	-	0.09	_	-	0.85	0.23 (0.08 to 0.38)	0.99 (0.48 to 1.50)	1.4×10 ⁻⁴	0.23 (0.11 to 0.35)	0.06 (-0.06 to 0.18)	0.13 (0.01 to 0.24)	
BA + BI	327	_	_	0.32	_	_	0.63	0.17 (0.02 to 0.32)	0.78 (0.27 to 1.29)	0.003	0.18 (0.06 to 0.30)	0.01 (-0.11 to 0.13)	0.08 (-0.04 to 0.19)	
Self-check [reference]	328	_	_	0.18	_	-	0.31	0.08 (-0.06 to 0.23)	0.45 (-0.05 to 0.96)	0.080	0.10 (-0.01 to 0.22)	-0.07 (-0.18 to 0.05)	_	
Health information [reference]	328	_	_	0.15	_	-	0.59	0.16 (0.01 to 0.31)	0.73 (0.23 to 1.24)	0.005	0.17 (0.05 to 0.29)	_	0.07 (-0.05 to 0.18)	
Delayed treatment [reference]	328	-	-	0.19	-	-	-0.15	-0.04 (-0.19 to 0.11)	_	_	_	-0.17 (-0.29 to -0.05)	-0.10 (-0.22 to 0.01)	

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression model included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. P-values are two-sided and no adjustments are made for multiple comparisons. † Minus value indicates depression reduction. () indicates 95% confidence intervals. ‡ Effect sizes (standardised mean differences) were calculated using the observed baseline SD for within-group change scores and the observed week 6 SD for group differences. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, DT: Delayed treatment, HI: Health information, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving, SC: Self-check.

Extended Data Table 6 | Time spent on the program (minutes)

			Acut	e intervention pl	nase		Follow-up phase					
	Introduction	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Epilogue	Weeks 6-10*	Weeks 10-14	Weeks 14-18	Weeks 18-22	Weeks 22-26
DA	20.7	40.0	20.2	13.3	13.5	15.0	9.2	22.1	3.6	1.7	1.5	1.3
DA	(11.9, 31.1)	(25.3, 63.0)	(10.9, 35.8)	(8.1, 26.6)	(8.0, 25.0)	(8.2, 27.7)	(3.0, 22.0)	(11.8, 41.1)	(1.3, 11.8)	(1.0, 5.2)	(0.9, 4.1)	(0.8, 5.8)
CP	23.0	45.0	20.0	14.3	19.3	23.3	11.9	19.7	2.1	1.4	1.2	1.2
Ch	(14.1, 35.4)	(28.3, 74.8)	(10.8, 32.3)	(8.4, 24.4)	(11.0, 32.2)	(14.8, 41.7)	(3.4, 24.3)	(10.8, 40.3)	(1.1, 8.3)	(0.8, 3.1)	(0.8, 2.7)	(0.8, 2.9)
DC	21.3	46.4	16.0	15.0	20.1	11.4	8.6	15.3	1.9	1.4	1.2	1.2
гJ	(12.8, 33.0)	(27.9, 70.2)	(8.8, 32.4)	(8.6, 27.9)	(11.5, 34.1)	(6.1, 23.3)	(3.4, 18.4)	(8.9, 26.1)	(1.0, 6.3)	(0.9, 3.0)	(0.8, 2.4)	(0.8, 3.0)
AT	21.3	43.9	23.3	10.9	16.6	9.8	11.2	17.0	1.7	1.3	1.3	1.1
AI	(13.1, 31.8)	(28.5, 68.9)	(15.4, 34.7)	(6.3, 19.4)	(9.4, 30.6)	(5.4, 18.7)	(3.9, 25.9)	(10.1, 30.3)	(1.0, 6.2)	(0.9, 2.6)	(0.8, 2.1)	(0.8, 2.1)
DI.	20.8	59.2	16.6	17.2	14.3	18.3	6.8	34.5	19.9	15.3	14.9	11.4
DI	(12.8, 32.2)	(39.0, 87.0)	(10.9, 30.0)	(10.7, 30.0)	(9.2, 26.2)	(11.3, 31.7)	(2.1, 14.8)	(18.9, 60.2)	(4.0, 35.8)	(2.5, 33.3)	(1.6, 33.2)	(1.6, 31.5)
PA+CP	21.3	36.9	27.2	40.3	29.0	18.0	9.4	20.8	2.1	1.5	1.3	1.2
DATCH	(14.4, 30.9)	(23.5, 60.0)	(15.1, 48.2)	(21.8, 63.4)	(17.2, 49.3)	(8.8, 30.9)	(2.9, 19.2)	(12.2, 42.8)	(1.0, 10.2)	(0.9, 5.3)	(0.8, 3.4)	(0.8, 3.2)
DA L DC	19.6	38.8	24.7	35.2	41.4	17.9	11.0	22.6	2.8	1.4	1.3	1.2
DATPO	(12.3, 30.9)	(23.0, 62.7)	(14.4, 42.2)	(20.3, 56.2)	(21.9, 67.6)	(9.7, 32.5)	(5.1, 19.4)	(11.4, 43.4)	(1.1, 11.3)	(0.9, 5.5)	(0.8, 3.0)	(0.8, 2.7)
DALAT	20.6	36.3	27.9	36.6	38.6	16.5	9.3	24.0	2.2	1.4	1.3	1.2
DATAI	(12.7, 30.7)	(22.7, 55.2)	(15.7, 47.4)	(19.4, 57.2)	(23.7, 61.6)	(8.9, 27.3)	(3.1, 22.8)	(10.6, 40.5)	(1.0, 9.2)	(0.8, 3.7)	(0.8, 2.8)	(0.7, 2.7)
DA + DI	20.7	41.7	27.1	56.3	26.5	24.8	6.3	40.9	12.3	7.0	4.2	4.7
DATDI	(11.9, 30.3)	(24.6, 61.4)	(15.0, 47.2)	(36.4, 83.8)	(14.6, 44.0)	(14.0, 42.1)	(2.6, 14.5)	(20.2, 73.8)	(2.5, 34.7)	(1.5, 26.6)	(1.0, 24.4)	(1.0, 27.2)
Colf obcoli							7.1	12.3	1.5	1.3	1.2	1.2
Sell-Check	-	-	-	-	-	-	(2.7, 13.1)	(7.7, 19.5)	(1.0, 3.5)	(0.8, 2.3)	(0.8, 2.0)	(0.8, 3.0)
Lloolth information	12.0	14.9	3.8	8.0			8.3	14.4	1.7	1.4	1.4	1.3
Health mormation	(5.9, 20.6)	(7.0, 26.5)	(2.1, 7.0)	(5.1, 14.7)	-	-	(3.7, 18.2)	(9.2, 23.5)	(1.1, 3.3)	(0.9, 2.4)	(0.9, 2.3)	(0.9, 2.3)
Delayed treatment							6.7	40.2	25.5	7.1	2.5	2.0
Delayeu treatment	-	-	-	-	-	-	(1.8, 13.7)	(16.4, 85.0)**	(8.4, 55.4)**	(1.9, 24.1)	(1.2, 12.4)	(1.1, 7.1)

Medians and interquartile ranges in parentheses. *The numbers for Weeks 6-10 may include the minutes spent after Week 6 but before the program completion, because we allowed up to 2 weeks before they could finish the program and respond to the Week 6 questionnaire. ** The Delayed treatment group engaged in the program during these eight weeks. AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, PS: Problem solving.

Extended Data Table 7 | Adherence

	Completed	all Chapters	Completed at leas	st up to Chapter 4
	By week 10	By week 30	By week 10	By week 30
BA	222 (68%)	272 (83%)	277 (85%)	291 (89%)
CR	166 (51%)	215 (66%)	252 (77%)	270 (82%)
PS	162 (49%)	225 (69%)	216 (66%)	238 (73%)
AT	198 (60%)	258 (79%)	244 (74%)	267 (81%)
BI	275 (84%)	301 (92%)	305 (93%)	309 (94%)
BA+CR	193 (59%)	253 (77%)	248 (75%)	270 (82%)
BA+PS	193 (59%)	252 (77%)	243 (74%)	266 (81%)
BA+AT	206 (63%)	254 (77%)	250 (76%)	264 (81%)
BA+BI	245 (75%)	287 (88%)	279 (85%)	294 (90%)
Self-check	229 (70%)	271 (83%)	-	-
Health information	294 (90%)	315 (96%)	320 (98%)	323 (99%)
Delayed treatment	296 (90%)	311 (95%)	-	-
Average completion rate of active interventions	63.0%	78.5%	78.4%	83.6%

AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, PS: Problem solving.

Extended Data Table 8 | Combined analysis using the mixed-model repeated measures analysis (MMRM) for PHQ-9 score at Week 26

Arm	N (26W)	N (26W)	Estimated least squares mean change scores of the PHQ-9	Effect size for Baseline-Week 26	Estimated differences of change scores (vs HI)		Effect size for differences of change scores [‡] (95% CI)	
		(2011)	at Week 26† (95% CI)	(95% CI)	(95% CI)	P-value	vs HI	vs SC
PA	207	202	-2.90	-1.06	-0.67	0.027	-0.19	-0.28
ВА	327	302	(-3.46 to -2.34)	(-1.26 to -0.85)	(-1.27 to -0.07)	0.027	(-0.36 to -0.02)	(-0.45 to -0.11)
<u>CP</u>	000	007	-2.87	-1.05	-0.64	0.005	-0.18	-0.27
CR	328	297	(-3.43 to -2.30)	(-1.25 to -0.84)	(-1.24 to -0.04)	0.035	(-0.35 to -0.01)	(-0.44 to -0.10)
DC.	000	000	-2.72	-0.99	-0.49	0.110	-0.14	-0.22
25	328	289	(-3.28 to -2.15)	(-1.20 to -0.78)	(-1.09 to 0.11)	0.112	(-0.31 to 0.03)	(-0.40 to -0.05)
A.T.	328	007	-2.76	-1.00	-0.53	0.000	-0.15	-0.24
AI		297	(-3.32 to -2.19)	(-1.21 to -0.80)	(-1.12 to 0.07)	0.082	(-0.32 to 0.02)	(-0.41 to -0.06)
			-3.12	-1.14	-0.90		-0.26	-0.34
ВІ	328	291	(-3.69 to -2.56)	(-1.34 to -0.93)	(-1.49 to -0.30)	0.003	(-0.43 to -0.09)	(-0.51 to -0.17)
			-3.32	-1.21	-1.09	3.6×10 ⁻⁴	-0.31	-0.40
BA + CR	329	290	(-3.89 to -2.75)	(-1.42 to -1.00)	(-1.69 to -0.49)		(-0.48 to -0.14)	(-0.57 to -0.23)
	329		-3.29	-1.20	-1.06		-0.30	-0.39
BA + PS		293	(-3.85 to -2.73)	(-1.40 to -0.99)	(-1.65 to -0.46)	0.001	(-0.47 to -0.13)	(-0.56 to -0.22)
			-3.15	-1.15	-0.92		-0.26	-0.35
BA + AI	328	281	(-3.72 to -2.58)	(-1.36 to -0.94)	(-1.52 to -0.32)	0.003	(-0.44 to -0.09)	(-0.52 to -0.18)
			-2.88	-1.05	-0.65		-0.19	-0.27
BA + BI	327	286	(-3.45 to -2.32)	(-1.26 to -0.84)	(-1.25 to -0.05)	0.033	(-0.36 to -0.02)	(-0.44 to -0.10)
	e] 328		-1.94	-0.71	0.29		0.08	
Self-check [reference]		317	(-2.49 to -1.39)	(-0.91 to -0.51)	(-0.30 to 0.88)	0.330	(-0.09 to 0.25)	-
		315	-2.23	-0.81				-0.08
Health information [reference]	328		(-2.78 to -1.67)	(-1.01 to -0.61)	_	—	—	(-0.25 to 0.09)

The least squares mean changes and their differences of the PHQ-9 scores were estimated with the mixed-effects models for repeated-measures (MMRM). The regression model included fixed effects of treatment, visit (as categorical), and treatment-by-visit interactions, adjusted for baseline PHQ-9 scores, employment status, age and gender. P-values are two-sided and no adjustments are made for multiple comparisons. tMinus value indicates depression reduction. () indicates 95% confidence intervals. tEffect sizes (standardised mean differences) were calculated using the observed baseline SD for within-group change scores and the observed week 6 SD for group differences. The Delayed treatment group engaged in the program during after week 6, and is therefore excluded from the week 26 analyses. Single-skill intervention arms showed smaller changes from baseline (-2.88, 95%CU: -3.37 to -2.40) than two-skill interventions (-3.17, 95%CI: -3.66 to -2.68), resulting in between-group differences of -02.9 (-0.57 to -0.00) in PHQ-9 scores or effect size of -0.08 (-0.16 to -0.00, p=0.049). AT: Assertion training, BA: Behavioural activation, BI: Behaviour therapy for insomnia, CR: Cognitive restructuring, HI: Health information, PHQ-9: Patient Health Questionnaire-9, PS: Problem solving, SC: Self-check.

1

nature portfolio

Corresponding author(s): Toshi A. Furukawa

Last updated by author(s): Feb 16, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	firmed
	\boxtimes	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	\boxtimes	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
		The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.
	\boxtimes	A description of all covariates tested
	\boxtimes	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
		A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
		For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable.
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	\boxtimes	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	\boxtimes	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
		Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

Software and code

 Policy information about availability of computer code

 Data collection
 The Resilience Training App (Version 2.1)

 Data analysis
 R (Version 4.3.1), mmrm (Version 0.3.11), emmeans (Version 1.9.0) R code files used in the statistical analyses are available on GitHub at https://github.com/nomahi/RESiLIENT.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability

De-identified individual participant data and data dictionary will be made available 24 months after publication on UMIN-ICDR, an individual case data repository managed by the Japanese University hospital Medical Information Network (UMIN) Center (https://www.umin.ac.jp/icdr/index.html). Proposals with specific aims

Research involving human participants, their data, or biological material

Policy information about studies with <u>human participants or human data</u>. See also policy information about <u>sex, gender (identity/presentation)</u>, <u>and sexual orientation</u> and <u>race, ethnicity and racism</u>.

Reporting on sex and gender	Gender was self-reported. There were 51% males, 29% females, and 0.5% others in the study cohort. Gender was entered as a covariate in the main analyses.
Reporting on race, ethnicity, or other socially relevant groupings	We did not ask participants questions regarding race, ethnicity or other socially relevant groupings.
Population characteristics	Table 1 presents the baseline characteristics of the participants.
Recruitment	Participants were recruited in the following four fields. The potential participants accessed the internet web page for the trial to apply for the study, received the explanations about the study from the trial coordinators online remotely, and provided their informed consent by electronic signature. 1. Health insurance societies: There are several types of public health insurance schemes together constituting the universal healthcare in Japan. We collaborated with two large associations, namely, National Federation of Health Insurance Societies covering employees and their families in large corporations (ca 30 million insured), and Japan Health Insurance Association covering employees and their families of small-to-middle-sized corporations (ca 35 million insured). 2. Business companies and corporations 3. Community and local governments 4. Direct-to-consumer advertisements We asked interested, potential participants to log onto the web specifically prepared for the trial and fill in the screening questionnaire. Once we ascertained the minimum eligibility criteria (age, possession of smartphone, not receiving treatment from mental health professionals), we invited them to the online informed consent session on Zoom in which we explained all the details of the study and its procedures. Once individuals have agreed and signed their digital consent forms, we provided the ID and password for the smartphone app to be downloaded either from App Store or from Google Play. Thus the participants were self-selected. This does not affect the internal validity of the study findings because all the comparisons were randomised. However, this may affect the external validity of the study findings. Thus in the limitations section of the main text, we discussed "Lastly, it is unclear whether our findings about the five skills are limited to our smartphone app only for adults with subthreshold depression and may not be generalisable to similarly administered CBT skills by other apps or by human therapists, extrapolate to the

Ethics oversight

The study was approved by Kyoto University Ethics Committee (C1556) on 3 March 2022.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

🔀 Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The RESiLIENT (Resilience Enhancement with Smartphone in Living ENvironemTs) trial is a master protocol trial involving four 2x2 factorial trials. This is a quantitative study.
Research sample	Participants were recruited in the following four fields across Japan. The potential participants accessed the internet web page for the trial to apply for the study, received the explanations about the study from the trial coordinators online remotely, and provided their informed consent by electronic signature. 1. Health insurance societies: There are several types of public health insurance schemes together constituting the universal healthcare in Japan. We collaborated with two large associations, namely, National Federation of Health Insurance Societies covering employees and their families in large corporations (ca 30 million insured), and Japan Health Insurance Association covering employees and their families of small-to-middle-sized corporations (ca 35 million insured). 2. Business companies and corporations 3. Community and local governments 4. Direct-to-consumer advertisements

	The participants were mainly in their 30s, 40s and 50s; 51% were men; most were married and employed. Their mean baseline PHQ-9 score was 8.1 (SD: 2.7); 17% had suspected problematic alcohol use, 29% had treatment history for mental health, and 33% had some physical comorbidities. (More details available in Table 1 of the main text.)
	They were intended to represent the general population in Japan as we used various access routes that we could think of to have access to as many people across Japan as possible. However, there is no guarantee that they are representative of the whole Japanese population.
Sampling strategy	The RESiLIENT trial was powered for each factorial trial for its primary outcome of the PHQ-9 at week 6 among those with the baseline PHQ-9 scores of 5 or more. The Fun the Learn, Act and Think through Technology trial (Mantani 2017) has shown a standardised mean difference of 0.3–0.4 for the earlier version of the Resilience Training App using the combination of BA+CR over the waiting list control among patients with drug-refractory depression. In order to detect a standardised mean difference of 0.2 at alpha=0.05 and beta=0.10, each factorial trial in the current meater protocol would need in total 10.2 a patients.
	Altogether the RESiLIENT trial will require 264×12=3168 participants with baseline PHQ-9 scores of 5 or more. The Healthy Campus Trial (HCT) had 5% dropout rates for its eighth week outcome. We may expect somewhat higher dropout rates in the RESiLIENT trial, which targets the general adult population in the community. Assuming a 10% dropout rate, the required sample size was set to be 3520.
Data collection	All the data were collected on the smartphone app developed for this study, The Resilience Training App (Version 2.1). The participants filled in the questionnaires on their own, with no researcher present either at the place or remotely. The research staff at the central office were not blind to the experimental condition or the study hypothesis.
Timing	Of 34123 adults assessed for eligibility, 5364 were judged eligible and provided informed consent, and were randomised to one of the 12 intervention or control arms between September 2022 and February 2024.
Data exclusions	Two later withdrew their consent, and 1426 had baseline PHQ-9 scores of 4 or less.
Non-participation	Across interventions, between 64% and 92% of the participants, or 78% on average, completed all the intervention programs within eight weeks, which was the time limit for the week 6 assessment. The proportions of the participants completing the assessments were 95.3% (3751/3936) at week 3 and 97.2% (3824/3936) at week 6 (Extended Data. Table 3.)
Randomization	After the participants completed all the questionnaires, the app automatically randomised the participants to one of the 12 intervention or control arms according to their employment status (yes vs no) using the pre-installed permuted block randomisation sequence, prepared by an independent statistician. The block size was known only to the statistician and allocation was concealed from any study personnel involved in participant recruitment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a Involve	d in the study	n/a	Involved in the study
Ant	ibodies	\boxtimes	ChIP-seq
Euk	aryotic cell lines	\boxtimes	Flow cytometry
Pala	aeontology and archaeology	\boxtimes	MRI-based neuroimaging
Anii	mals and other organisms		
Clin	ical data		
Dua 🗌	I use research of concern		
Plar	nts		

Clinical data

Policy information about <u>clinical studies</u> All manuscripts should comply with the ICMJE <u>guidelines for publication of clinical research</u> and a completed <u>CONSORT checklist</u> must be included with all submissions. Clinical trial registration UMIN Clinical Trials Registry (UMIN000047124) on 1 Aug, 2022

Study protocol	Furukawa TA, Tajika A, Sakata M, Luo Y, Toyomoto R, Horikoshi M, Akechi T, Kawakami N, Nakayama T, Kondo N, Fukuma S, Noma H, Christensen H, Kessler RC, Cuijpers P & Wason JMS (2023) Four 2x2 factorial trials of smartphone CBT to reduce subthreshold depression and to prevent new depressive episodes among adults in the community-RESiLIENT trial (Resilience Enhancement with Smartphone in Llving ENvironmenTs): a master protocol. BMJ Open, 13, e067850. 10.1136/bmjopen-2022-067850
Data collection	The recruitment of participants took place between 5 September 2022 and 21 February 2024. The date of the last-participant-out for the week 6 data was 19 April 2024, that for the week 26 data was 3 September 2024. The participants filled in the questionnaires

within the app (The Resilience Training App, Version 2.1) using their own smartphone, with no researcher present either at the place or remotely.

Outcomes	Primary outcome The primary outcome is the change in the PHQ-9 from baseline to week 6.
	Secondary outcome The secondary outcomes include: 1. Changes in PHQ-9 from baseline to weeks 1, 2, 3, 4, and 5. 2. Changes in the GAD-7 51 from baseline to weeks 3 and 6. 3. Changes in the ISI 63 from baseline to weeks 3 and 6. 4. Changes in the SWEMWBS 62 from baseline to weeks 3 and 6. Anxiety usually coexists with depression and we aimed to examine what effects the intervention components, while mainly targeting depression, may have on anxiety. We also aimed to monitor broad psychopathology of common mental disorders encompassing depression, anxiety and insomnia as well as positive mental health which are deteriorated in common mental disorders. Post-hoc additional analyses We added two post-hoc analyses up to week 26 to inform and facilitate the interpretations of the primary analysis results. All the outcomes were collected via Resilience Training App.

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor
Authentication	was applied. Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.