

Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials

Jake Linardon^{1,2}, John Torous³, Joseph Firth^{4,5}, Pim Cuijpers^{6,7}, Mariel Messer¹, Matthew Fuller-Tyszkiewicz^{1,2}

¹School of Psychology, Deakin University, Geelong, VIC, Australia; ²Center for Social and Early Emotional Development, Deakin University, Burwood, VIC, Australia; ³Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; ⁴Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; ⁵Greater Manchester Mental Health NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; ⁶Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; ⁷International Institute for Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania

The mental health care available for depression and anxiety has recently undergone a major technological revolution, with growing interest towards the potential of smartphone apps as a scalable tool to treat these conditions. Since the last comprehensive meta-analysis in 2019 established positive yet variable effects of apps on depressive and anxiety symptoms, more than 100 new randomized controlled trials (RCTs) have been carried out. We conducted an updated meta-analysis with the objectives of providing more precise estimates of effects, quantifying generalizability from this evidence base, and understanding whether major app and trial characteristics moderate effect sizes. We included 176 RCTs that aimed to treat depressive or anxiety symptoms. Apps had overall significant although small effects on symptoms of depression ($N=33,567$, $g=0.28$, $p<0.001$; number needed to treat, $NNT=11.5$) and generalized anxiety ($N=22,394$, $g=0.26$, $p<0.001$, $NNT=12.4$) as compared to control groups. These effects were robust at different follow-ups and after removing small sample and higher risk of bias trials. There was less variability in outcome scores at post-test in app compared to control conditions (ratio of variance, $RoV=-0.14$, 95% CI: -0.24 to -0.05 for depressive symptoms; $RoV=-0.21$, 95% CI: -0.31 to -0.12 for generalized anxiety symptoms). Effect sizes for depression were significantly larger when apps incorporated cognitive behavioral therapy (CBT) features or included chatbot technology. Effect sizes for anxiety were significantly larger when trials had generalized anxiety as a primary target and administered a CBT app or an app with mood monitoring features. We found evidence of moderate effects of apps on social anxiety ($g=0.52$) and obsessive-compulsive ($g=0.51$) symptoms, a small effect on post-traumatic stress symptoms ($g=0.12$), a large effect on acrophobia symptoms ($g=0.90$), and a non-significant negative effect on panic symptoms ($g=-0.12$), although these results should be considered with caution, because most trials had high risk of bias and were based on small sample sizes. We conclude that apps have overall small but significant effects on symptoms of depression and generalized anxiety, and that specific features of apps – such as CBT or mood monitoring features and chatbot technology – are associated with larger effect sizes.

Key words: Smartphone apps, depression, generalized anxiety, social anxiety, post-traumatic stress, panic, cognitive behavioral therapy, mood monitoring, chatbot technology

(*World Psychiatry* 2024;23:139–149)

Depressive and anxiety disorders are common mental health conditions associated with significant disease burden, profound economic costs, premature mortality, and severe quality of life impairments for affected individuals and their relatives^{1,2}. Widely accessible treatments are required to reduce these impacts. Both psychotherapies and pharmacotherapies can effectively treat symptoms of depression and anxiety^{3–7}. However, there are many barriers that prevent people from accessing these traditional forms of treatment, including limited availability of trained psychiatrists and psychologists, high cost of treatment, stigma associated with help-seeking, and low mental health literacy^{8,9}.

Over the past two decades, the mental health care available for depression and anxiety has undergone a major technological revolution. Empirically validated components of psychological interventions that were once delivered solely in-person have now been translated for delivery via low-cost, private and scalable digital tools^{10,11}.

Smartphone applications (“apps”) are one form of digital treatment delivery which is receiving substantial attention. Smartphones are among the most rapidly adopted technological innovations in recent history. Over 6.5 billion people own a smartphone, that is typically checked multiple times per day and always kept within arm’s reach¹². Treatment content delivered via an app can thus be accessed anytime and anywhere, enabling users

to practice those critical therapeutic skills that are necessary for preventing the onset or escalation of symptoms in moments of need⁸. Digital monitoring systems and complex machine learning algorithms also enable treatment content to be regularly updated and personalized to the needs of the user based on data collected passively (e.g., global positioning system coordinates to infer social determinants of health) and actively (e.g., symptom tracking)^{13,14}.

The potential of apps to treat depressive and anxiety symptoms is attracting an increasing interest among patients, clinicians, technology companies, and health care regulators. However, there are risks associated with depression and anxiety apps, such as privacy violations; possible easy access to ineffective, inaccurate or potentially harmful content^{15,16}; low rates of engagement^{17,18}, and exclusion of the potential therapeutic ingredient represented by the personal relationship between a clinician and a patient. All this, coupled with the fact that mental health apps for depression and anxiety are some of the most widely publicly offered and downloaded categories of health apps^{19,20}, generates the duty to provide the public with up-to-date information on the evidence base supporting their use²¹.

In a 2019 meta-analysis, Linardon et al¹³ found apps to outperform control conditions in reducing symptoms of depression ($g=0.28$) and generalized anxiety ($g=0.30$) based on 41 and 28 trials, respectively. This meta-analysis also found early evidence

for the efficacy of apps on social anxiety – but not post-traumatic stress or panic – symptoms (≤ 6 trials). Heterogeneity in efficacy was noted, although few robust effect modifiers were identified, perhaps reflecting the relatively low numbers of trials available²² and the focus on univariate moderator effects rather than more complex multivariate moderator models.

Research testing mental health apps on depressive and anxiety symptoms is growing exponentially, offering greater opportunity to explore whether recent innovations in digital health have promoted improved efficacy, and to examine the individual and combined effects of moderators on treatment efficacy. Since 2019, there have been more than 100 randomized controlled trials (RCTs) published, some of which include large sample sizes, use credible comparison conditions, and have a lower risk of bias. Each of these elements had been raised as critical features absent in this field according to the conclusions drawn by the authors of earlier meta-analyses^{13,21,23}. Furthermore, the large number of available trials now enables more precise, complex and adequately powered analysis of those app and trial characteristics that may be associated with effect sizes.

In light of the limitations of past reviews, and with interest in research on mental health apps further expanding, we conducted an updated meta-analysis testing the effects of mental health apps on symptoms of depression and anxiety. In addition to typical pooling of mean differences between intervention and control conditions, we also explored group differences in variability of outcomes, as a means to gauge the potential generalizability of the effects of these apps. As argued recently^{24,25}, if an intervention has variable effect on participants, this may be observed through greater variability around the post-test mean for the intervention group relative to the control group. Finally, in light of evident heterogeneity of effect sizes from prior reviews^{13,21}, we attempted to identify moderators (specifically, major characteristics of the app, study population, and trial design) that may account for larger or smaller effect sizes than the pooled average.

Furthermore, this meta-analysis aimed to move beyond examination of individual moderators, to also – for the first time – evaluate potential combinations and interactions among the pre-selected moderator variables²⁶. This will help shed new light on the contexts in which specific intervention components may be most effective, and further characterize subgroups of individuals who respond particularly well to mental health apps.

METHODS

Identification and selection of trials

This review was conducted in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines²⁷, and adhered to a pre-registered protocol (CRD42023437664). We searched (last updated June 2023) the Medline, PsycINFO, Web of Science, and ProQuest Database for Dissertations online databases combining key terms related to smartphones, RCTs, and anxiety or depression (see supplementary

information for the full search strategy). We also hand-searched through relevant reviews^{13,14,21} and reference lists of included studies to identify any potentially eligible trials not captured in the primary search.

We included RCTs that tested the effects of a stand-alone, smartphone-based, mental health app against a control condition (e.g., waitlist, placebo, information resources) or an active comparison (e.g., face-to-face treatment) for symptoms of depression or anxiety. Trials that conducted one psychoeducational or information session prior to the delivery of the app program were eligible for inclusion. Blended web and app-based programs were excluded, as were apps not focused – solely or in part – on targeting mental health (e.g., weight loss or diet apps). Text-message only interventions were excluded. Adjunctive treatments were also excluded, such as when apps were incorporated within a broader face-to-face psychotherapy program. However, trial arms comprised of a mental health app plus usual care were included, as long as the usual care component did not consist of a structured psychological treatment program (e.g., trials of patients with a medical condition continuing their usual care were permitted). Published and unpublished trials were eligible for inclusion. If a study did not include data for effect size calculation, the authors were contacted, and the study was excluded if they failed to provide the data.

Quality assessment and data extraction

We used criteria from the Cochrane Collaboration Risk of Bias tool²⁸ to assess for risk of bias. These criteria include random sequence generation, allocation concealment, blinding of participants or personnel, blinding of outcome assessment, and completeness of outcome data. Each domain was rated as high risk, low risk, or unclear. Selection bias was rated as low risk if there was a random component in the allocation sequence generation. Allocation concealment was rated as low risk when a clear method that prevented foreseeing group allocation before or during enrolment was explicitly stated. Blinding of participants was rated as low risk when the trial incorporated a comparison condition that prevented participants from knowing whether they were assigned to the experimental or control condition (e.g., a placebo app or an intervention intended to be therapeutic). Blinding of outcome assessors was rated as low risk if proper measures were taken to conceal participants' group membership, or if the outcome measures were self-reported, which does not involve direct contact with the researcher. Completeness of outcome data was rated as low risk if the trial authors included all randomized participants in their analyses (i.e., they adhered to the intention-to-treat principle).

We also extracted several characteristics pertaining to the study (year, author, sample size), participants (target group, selection criteria), app intervention (orientation, primary target, key features, prescription of human guidance), comparison (type of control condition), and outcome assessment (tool used, length assessed, primary vs. secondary). Two researchers performed data extraction, and any disagreement was resolved through consensus.

Meta-analysis

Analyses were conducted using Comprehensive Meta-Analysis Version 3.0²⁹ for effect size estimates of between-group mean differences and univariate subgroup analyses, and R for comparing variability between intervention and control groups and for probing interactions among moderators. For each comparison of means between the app intervention and the control condition, the effect size was calculated by dividing the difference between the two group means at post-test by the pooled standard deviation. We reported Hedges' *g* over Cohen's *d* to correct for small sample bias³⁰. If means and standard deviations were not reported, effect sizes were calculated using change scores or other reported statistics (*t* or *p* values for group comparisons). To calculate a pooled effect size, each study's effect size was weighted by its inverse variance. If multiple measures of a given outcome variable were used, the mean of the effect sizes for each measure within the study was calculated, before the effect sizes were pooled²⁹. A positive *g* indicates that the app condition achieved higher symptom reduction than the comparison condition. Effect sizes of 0.8 were interpreted as large, while effect sizes of 0.5 as moderate, and effect sizes of 0.2 as small³¹. In the protocol we had stated that we would conduct meta-analyses on rates of remission, recovery and reliable change. However, because these outcomes were rarely (<10% of eligible trials) and inconsistently reported, we were not able to conduct these analyses.

Meta-analysis of differences in variability at post-test for app and control groups was undertaken, as it provides indication of whether intervention effects are reasonably uniform. If variability estimates for the app group are comparable to, or smaller than, those found for the control group, it is suggested that the intervention may have good generalizability potential. In contrast, greater variability for the app group suggests that effects may be limited to a subset of participants²⁴. We conducted this comparison of variability estimates by deriving a log-transformed estimate of the ratio of app group variance to control group variance in post-test outcomes²⁴. Alternate ways of quantifying differences in variability²⁴ were tested for robustness of our initial results. The significance and direction of differences in variance between groups remained the same regardless of operationalization used.

We also conducted several other sensitivity analyses to assess whether the above main outcomes were robust. We re-calculated the pooled effects when restricting the analyses to: a) lower risk of bias trials (defined as meeting 4 or 5 of the quality criteria); b) larger sample trials (defined as 75 or more randomized participants per condition); c) trials delivering an app that was explicitly designed to address depression or anxiety symptoms, or when depression or anxiety was declared as the primary outcome; d) the smallest and largest effect in each study, if multiple conditions were used (to maintain statistical independence); and e) different post-test lengths (1-4 weeks, 5-12 weeks, or 13 or more weeks). We also pooled effects while excluding outliers using the non-overlapping confidence interval (CI) approach, in which a study is defined as an outlier when the 95% CI of the

effect size does not overlap with the 95% CI of the pooled effect size³². Small-study bias was also examined through the trim-and-fill method³³.

Between-group effect size estimates were supplemented with estimates of the number needed to treat (NNT), to convey the practical impact of the weighted mean for intervention effects. NNT indicates the number of additional participants in the intervention group who would need to be treated in order to observe one participant who shows positive symptom change relative to the control group³⁴.

We also calculated the weighted average dropout rate from app conditions of included trials. This was defined as the proportion of participants assigned to the app condition who did not complete the post-test assessment, divided by the total number of participants randomized to that condition. Event rates calculated through Comprehensive Meta-Analysis were converted to percentages for ease of readability.

Since we expected considerable heterogeneity among the studies, random effects models were employed for all analyses²⁹. Heterogeneity was examined by calculating the *I*² statistic, which quantifies heterogeneity revealed by the *Q* statistic and reports how much overall variance (0-100%) is attributed to between-study variance³⁵. We conducted a series of univariate subgroup analyses, examining the effects of the intervention according to major characteristics of participants, app features, and trials (see also supplementary information). Subgroup analyses were conducted under a mixed effects model²⁹.

Finally, recognizing that determinants of effect size may interact in complex ways that are not adequately captured through univariate subgroup analysis techniques, we evaluated interactive effects of moderators on effect estimates through meta-CART²⁶. This takes a list of potential moderators and seeks to partition scores on a key outcome variable (in this case, effect sizes from each trial) according to combinations of these moderators that maximize between-group differences in the outcome whilst minimizing within-group variance. This process of partitioning continues until the set of effect sizes cannot be significantly improved through further splitting into subgroups. We used 10-fold cross-validation and random-effects modelling, given expectations of multiple sources of variability per effect size in the analysis. Past research³⁶ suggests that meta-CART is well powered to detect interaction among potential moderators for cases where there are at least 80 estimates in the sample – a condition met in the present review.

RESULTS

Study characteristics

Figure 1 presents a flow chart of the literature search. A total of 176 RCTs from 174 papers met full inclusion criteria (see supplementary information for the characteristics of individual studies). More than two-thirds of eligible trials (67%) were conducted between 2020 and 2023. Many trials (43%) recruited an

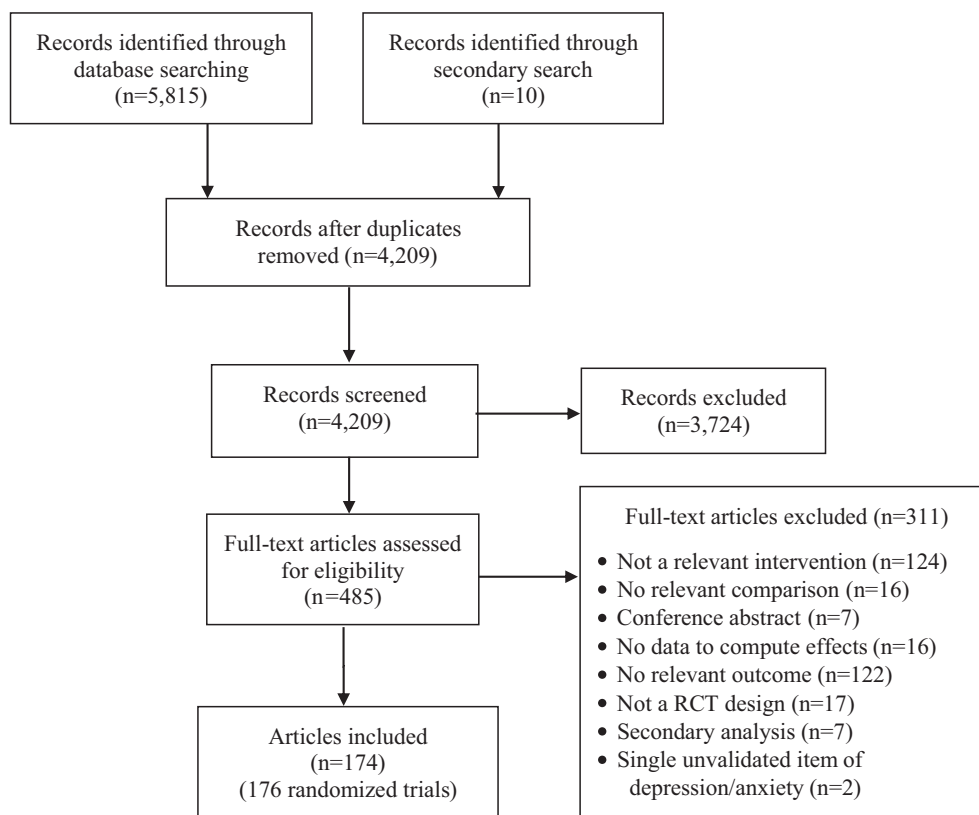


Figure 1 PRISMA flow chart. RCT – randomized controlled trial

unselected convenience sample, while a smaller number recruited people with depression or generalized anxiety, either meeting diagnostic criteria (6%) or scoring above a cut-off on a validated measure (26%). Fewer trials recruited participants with post-traumatic stress, social anxiety, obsessive-compulsive or panic symptoms (10%).

Nearly half of the apps delivered (48%) were based on cognitive behavioral therapy (CBT) principles; fewer apps were based on mindfulness (21%) or cognitive training (10%). A third of the apps (34%) had mood monitoring features, while only 5% incorporated chatbot technology. Human guidance was offered in 14% of the apps delivered.

Most trials delivered an inactive control (60%), comprised of a waitlist, assessment only, or information resources. Fewer trials (23%) delivered a placebo control (e.g., non-therapeutic app, ecological momentary assessments) that attempted to control for participant time, attention or expectations. Care as usual was delivered in 11% of trials. Only ten trials used an active psychological comparison, such as face-to-face treatment sessions or a web-based program. Most trials employed a short-term follow-up of 1-4 (56%) or 5-12 (40%) weeks.

Risk of bias also varied. All trials used a self-report measure of depression or anxiety; 70% met criteria for adequate sequence generation; 25% met criteria for adequate allocation concealment; 26% met criteria for sufficiently blinding participants to study conditions, and 62% reported use of intention-to-treat analyses. Few trials (6%) met all five criteria, 21% met four criteria,

32% met three criteria, 31% met two criteria, and 8% only met one criterion.

Effects on depressive symptoms

Apps versus control conditions

The pooled effect size for the 181 comparisons between apps (N=16,569) and control conditions (N=17,007) on symptoms of depression was $g=0.28$ (95% CI: 0.23-0.33, $p<0.001$; $I^2=72%$, 95% CI: 67-75), corresponding to an NNT of 11.5 (see Table 1). The pooled estimate for ratio of variance (RoV) analyses was -0.14 (95% CI: -0.24 to -0.05), indicating less variance in post-test outcome scores for the app intervention group relative to control group. However, heterogeneity was high for differences in variance ($I^2=78%$), suggesting variable effects of app interventions.

Sensitivity analyses supported the main findings (see Table 1). The pooled effect size was similar when restricting the analyses to lower risk of bias and larger sample trials; when adjusting for small-study bias according to the trim-and-fill procedure; when limiting to trials where depression was the primary intervention target or outcome; and at different follow-up lengths. When excluding outliers, the pooled effect size was also similar, and heterogeneity substantially decreased. Across sensitivity analyses, heterogeneity was high for differences in variance, further suggesting variable effects of app interventions (see supplementary information).

Table 1 Meta-analyses on the effects of apps on symptoms of depression and generalized anxiety

Analysis	Depressive symptoms					Generalized anxiety symptoms				
	n	g (95% CI)	p	I ²	NNT	n	g (95% CI)	p	I ²	NNT
Apps vs. control conditions	181	0.28 (0.23-0.33)	<0.001	72%	11.5	150	0.26 (0.21-0.31)	<0.001	64%	12.4
Lower risk of bias trials only	48	0.32 (0.23-0.40)	<0.001	80%	9.9	35	0.25 (0.18-0.33)	<0.001	49%	12.9
Small sample trials removed	60	0.22 (0.15-0.28)	<0.001	79%	14.9	47	0.22 (0.16-0.28)	<0.001	67%	14.9
Primary intervention target or outcome	60	0.38 (0.28-0.48)	<0.001	84%	8.1	45	0.20 (0.13-0.28)	<0.001	53%	16.5
Outliers removed	147	0.25 (0.21-0.28)	<0.001	24%	12.9	131	0.23 (0.20-0.27)	<0.001	16%	14.2
One effect per study (smallest)	145	0.27 (0.22-0.32)	<0.001	73%	11.9	121	0.25 (0.20-0.31)	<0.001	69%	12.9
One effect per study (largest)	146	0.31 (0.25-0.35)	<0.001	75%	10.2	121	0.29 (0.24-0.35)	<0.001	68%	11.0
Trim-and-fill procedure	177	0.29 (0.24-0.33)	-	75%	11.0	148	0.26 (0.21-0.31)	-	64%	12.4
Follow-up duration										
1-4 weeks	102	0.23 (0.17-0.29)	<0.001	65%	14.2	95	0.21 (0.16-0.26)	<0.001	53%	15.6
5-12 weeks	76	0.35 (0.27-0.43)	<0.001	70%	8.9	54	0.34 (0.24-0.44)	<0.001	74%	9.2
≥13 weeks	3	0.29 (-0.17 to 0.76)	0.214	87%	11.0	1	0.29 (-0.13 to 0.73)	0.180	0%	11.0
Apps vs. active comparisons	8	-0.08 (-0.25 to 0.08)	0.340	0%	-	6	0.11 (-0.24 to 0.47)	0.537	64%	31.0
Face-to-face comparator only	5	-0.12 (-0.35 to 0.09)	0.257	0%	-	5	0.16 (-0.25 to 0.59)	0.441	65%	20.9
Web-based comparator only	3	-0.01 (-0.31 to 0.29)	0.962	16%	-	1	-0.11 (-0.52 to 0.29)	0.575	0%	-

n – number of comparisons, NNT – number needed to treat

Table 2 presents the results from the univariate subgroup analyses. Effect sizes for depression were significantly larger when trials used an inactive control group (relative to placebo or care as usual), studied a pre-selected sample (relative to an unselected sample), administered a CBT app (relative to a non-CBT app), and delivered an app that contained chatbot technology (relative to no chatbot technology).

We re-computed the univariate subgroup analyses only among trials where depression was the primary intervention target or outcome (see Table 3). In these exploratory analyses, the same subgroup effects emerged, with one exception: trials that delivered a mindfulness app produced significantly lower effect sizes than those that did not deliver a mindfulness app.

Meta-CART analyses identified five key trial features (pre-selected sample, sample size, control group, psychiatric diagnosis, and delivery of a cognitive training app) that characterized subgroups with higher or lower effect estimates for mean differences than for the sample overall. The sample of effect size estimates for depression was first split by whether the sample was pre-selected. Effect estimates from samples that were not pre-selected ($g=0.22$, 95% CI: 0.18-0.27, $n=136$) could not be split into further subgroups. Effect sizes for trials involving pre-selected samples were further split into subgroups reflecting: a) larger sample trials (≥ 75 per condition) plus placebo control groups ($g=0.06$, 95% CI: -0.14 to 0.25, $n=6$); b) larger sample trials plus inactive control groups ($g=0.54$, 95% CI: 0.35-0.74, $n=6$); c) smaller sample trials plus samples with a psychiatric diagnosis ($g=0.14$, 95% CI: -0.19 to 0.47, $n=4$); d) smaller sample trials plus samples without a psychiatric diagnosis plus apps that were not cognitive training-focused ($g=0.59$, 95% CI: -0.46 to 0.72, $n=26$), and e) smaller sample trials plus samples

without a psychiatric diagnosis plus cognitive training app ($g=1.16$, 95% CI: 0.73- 1.60, $n=3$) (see also supplementary information).

Apps versus active interventions

The pooled effect size for the eight comparisons between apps and active interventions was $g=-0.08$ (95% CI: -0.25 to 0.08, $p=0.340$). The effect size was $g=-0.12$ (95% CI: -0.35 to 0.09, $p=0.257$) when restricting the analyses to comparisons with face-to-face treatments and $g=-0.01$ (95% CI: -0.31 to 0.29, $p=0.962$) when restricting the analyses to comparisons with web-based interventions, although the number of studies in these analyses was low (see Table 1).

Effects on generalized anxiety symptoms

Apps versus control conditions

The pooled effect size for the 150 comparisons between apps ($N=10,972$) and control conditions ($N=11,422$) on symptoms of generalized anxiety was $g=0.26$ (95% CI: 0.21-0.31, $p<0.001$; $I^2=64%$, 95% CI: 57-69), corresponding to an NNT of 12.4 (see Table 1). The pooled RoV estimate was -0.21 (95% CI: -0.31 to -0.12), indicating less variance in post-test outcome scores for the app group relative to control group. However, heterogeneity was high for differences in variance ($I^2=75%$), suggesting variable efficacy of app interventions. Effects were similar across extensive sensitivity analyses reported in Table 1.

Table 2 Subgroup analyses on all available trials

Analysis	Depressive symptoms					Generalized anxiety symptoms				
	n	g (95% CI)	I ²	NNT	p	n	g (95% CI)	I ²	NNT	p
Control group					0.003					0.216
Inactive	112	0.33 (0.27-0.39)	68%	9.5		96	0.28 (0.23-0.33)	48%	11.4	
Placebo	49	0.19 (0.11-0.27)	70%	17.4		38	0.19 (0.11-0.28)	55%	17.4	
Care as usual	20	0.18 (0.07-0.27)	49%	18.4		16	0.32 (0.09-0.54)	88%	9.2	
Sample					<0.001					0.080
Pre-selected	45	0.52 (0.38-0.65)	88%	5.2		32	0.35 (0.23-0.46)	62%	8.9	
Unselected	136	0.21 (0.17-0.26)	56%	15.6		118	0.24 (0.18-0.29)	64%	13.5	
Psychiatric diagnosis					0.669					0.068
Yes	14	0.23 (0.05-0.40)	60%	14.2		10	0.41 (0.24-0.58)	35%	7.5	
No	167	0.28 (0.23-0.33)	73%	11.4		140	0.25 (0.20-0.30)	65%	12.9	
CBT app					0.003					0.147
Yes	86	0.35 (0.28-0.42)	79%	8.9		59	0.30 (0.24-0.36)	46%	10.6	
No	95	0.21 (0.15-0.27)	59%	15.6		91	0.23 (0.16-0.30)	69%	14.2	
Mindfulness app					0.549					0.258
Yes	43	0.26 (0.19-0.33)	45%	12.4		45	0.23 (0.15-0.30)	46%	14.2	
No	138	0.29 (0.23-0.34)	75%	11.0		105	0.28 (0.22-0.35)	68%	11.4	
Cognitive training app					0.238					0.370
Yes	21	0.18 (0.02-0.35)	66%	18.4		18	0.20 (0.07-0.33)	39%	16.5	
No	160	0.29 (0.24-0.34)	73%	11.0		132	0.26 (0.21-0.31)	65%	12.4	
Mood monitoring features					0.257					0.369
Yes	65	0.24 (0.19-0.30)	44%	13.5		51	0.23 (0.17-0.29)	33%	14.2	
No/Not reported	116	0.29 (0.23-0.36)	78%	11.0		99	0.27 (0.20-0.34)	71%	11.9	
Chatbot feature					0.009					0.258
Yes	12	0.53 (0.33-0.74)	61%	5.6		10	0.18 (0.06-0.31)	0%	18.4	
No/Not reported	169	0.26 (0.21-0.31)	72%	12.4		140	0.26 (0.21-0.32)	66%	12.4	
Human guidance offered					0.936					0.477
Yes	27	0.29 (0.13-0.46)	65%	11.0		18	0.37 (0.03-0.71)	88%	8.4	
No/Not reported	154	0.28 (0.23-0.32)	73%	11.4		132	0.24 (0.20-0.29)	52%	13.5	

n – number of comparisons, NNT – number needed to treat, CBT – cognitive behavioral therapy

In the univariate subgroup analyses of all available trials reporting generalized anxiety as an outcome, no significant moderation effects were found (see Table 2). However, when restricting these analyses to trials where generalized anxiety was the primary target, several univariate moderation effects emerged: trials that used an inactive control (relative to placebo or care as usual), pre-selected participants for generalized anxiety symptoms (relative to an unselected sample), administered a CBT app (relative to a non-CBT app), and delivered an app with mood monitoring features produced significantly larger effect sizes on generalized anxiety symptoms. In contrast, trials that delivered a mindfulness or cognitive training app produced significantly smaller effect sizes on gener-

alized anxiety symptoms (see Table 3).

Meta-CART analyses identified three key moderators (pre-selection, whether generalized anxiety was the primary target/outcome, and apps with mood monitoring features) that characterized subgroups with higher or lower effect estimates of mean differences than for the sample overall. The sample of effect size estimates was first split by whether the sample was pre-selected. For samples that were pre-selected, effect estimates were further split into whether the app included mood monitoring features (g=0.54, 95% CI: 0.36-0.27, n=12) or not (g=0.26, 95% CIs: 0.13,-0.39, n=20). For samples that were not pre-selected, effect estimates were split based on whether anxiety was the primary target/outcome

Table 3 *Post-hoc* subgroup analyses on trials where depression or generalized anxiety was the primary intervention target or outcome

Analysis	Depressive symptoms					Generalized anxiety symptoms				
	n	g (95% CI)	I ²	NNT	p	n	g (95% CI)	I ²	NNT	p
Control group					0.005					0.050
Inactive	29	0.56 (0.37-0.74)	82%	5.2		25	0.30 (0.18-0.41)	47%	10.6	
Placebo	21	0.30 (0.16-0.44)	82%	10.6		15	0.12 (0.02-0.21)	32%	28.3	
Care as usual	10	0.16 (0.01-0.31)	67%	20.9		5	0.11 (-0.10 to 0.33)	72%	31.0	
Sample					<0.001					<0.001
Pre-selected	40	0.55 (0.40-0.70)	88%	5.4		24	0.34 (0.22-0.47)	61%	9.2	
Unselected	20	0.12 (0.04-0.21)	35%	28.3		21	0.09 (0.03-0.15)	0%	38.2	
Psychiatric diagnosis					0.364					0.071
Yes	4	0.18 (-0.26 to 0.63)	67%	18.4		7	0.40 (0.16-0.64)	46%	7.7	
No	56	0.40 (0.29-0.50)	85%	7.7		38	0.17 (0.09-0.24)	46%	19.6	
CBT app					0.041					0.029
Yes	40	0.46 (0.32-0.58)	87%	6.6		22	0.28 (0.18-0.38)	41%	11.4	
No	20	0.24 (0.09-0.40)	78%	13.5		23	0.12 (0.01-0.23)	52%	28.3	
Mindfulness app					0.024					0.036
Yes	9	0.21 (0.07-0.35)	31%	15.6		9	0.07 (-0.06 to 0.21)	24%	49.5	
No	51	0.42 (0.31-0.53)	85%	7.3		36	0.24 (0.15-0.33)	56%	13.5	
Cognitive training app					0.218					0.016
Yes	4	0.85 (0.07-1.63)	83%	3.3		7	0.03 (-0.10 to 0.17)	52%	117.5	
No	56	0.35 (0.26-0.45)	83%	8.9		38	0.23 (0.15-0.31)	19%	14.2	
Mood monitoring features					0.327					0.033
Yes	26	0.33 (0.21-0.45)	65%	9.5		25	0.28 (0.18-0.37)	44%	11.4	
No/Not reported	34	0.42 (0.27-0.57)	88%	7.3		20	0.12 (0.01-0.22)	52%	28.3	
Chatbot feature					0.005					0.493
Yes	5	0.80 (0.50-1.10)	47%	3.5		5	0.21 (0.13-0.29)	0%	15.6	
No/Not reported	55	0.34 (0.25-0.44)	84%	9.2		40	0.14 (-0.04 to 0.33)	56%	24.1	
Human guidance offered					0.503					0.358
Yes	8	0.52 (0.11-0.92)	80%	5.7		3	0.45 (-0.10 to 1.01)	84%	6.7	
No/Not reported	52	0.37 (0.27-0.47)	85%	8.4		42	0.19 (0.11-0.26)	47%	17.4	

n – number of comparisons, NNT – number needed to treat, CBT – cognitive behavioral therapy

(g=0.08, 95% CI: -0.05 to 0.19, n=21) or not (g=0.28, 95% CI: 0.22-0.34, n=97) (see also supplementary information).

intervention (see Table 1).

Apps versus active interventions

The pooled effect size for the six comparisons between apps and active interventions was g=0.11 (95% CI: -0.24 to 0.47, p=0.537). The effect size was g=0.16 (95% CI: -0.25 to 0.59, p=0.441) when restricting the analyses to comparisons with face-to-face treatments, and g=-0.11 (95% CI: -0.52 to 0.29, p=0.575) when restricting the analyses to the one comparison with a web-based

Effects on specific anxiety symptoms

The effects of apps as compared to control conditions on specific anxiety symptoms are presented in Table 4. Subgroup analyses and analyses comparing apps to active interventions were not performed on these outcomes due to the limited number of available trials.

The pooled effect size for the 17 comparisons between apps (N=1,371) and control conditions (N=1,385) on post-traumatic

Table 4 Meta-analyses on the effects of apps on specific anxiety symptoms

Outcome	Analysis	n	g (95% CI)	p	I ²	NNT
Post-traumatic stress symptoms						
	Apps vs. control conditions	17	0.12 (0.03-0.21)	0.007	24%	28.3
	Lower risk of bias trials only	3	0.34 (0.11-0.57)	0.004	22%	9.2
	Small sample trials removed	5	0.11 (-0.04 to 0.25)	0.150	59%	31.0
	One effect per study (smallest)	15	0.13 (0.03-0.23)	0.008	29%	26.0
	One effect per study (largest)	15	0.14 (0.04-0.24)	0.005	30%	24.1
	Trim-and-fill procedure	14	0.13 (0.02-0.24)	-	-	26.0
	Post-traumatic stress primary target/outcome	12	0.12 (0.01-0.24)	0.039	44%	28.3
	Pre-selected for post-traumatic stress symptoms	10	0.14 (0.04-0.25)	0.006	29%	24.1
	CBT apps only	12	0.15 (0.02-0.29)	0.019	35%	22.4
Social anxiety symptoms						
	Apps vs. control conditions	10	0.52 (0.22-0.82)	0.001	75%	5.7
	Lower risk of bias trials only	1	0.10 (-0.16 to 0.37)	0.446	0%	34.3
	Small sample trials removed	1	0.10 (-0.16 to 0.37)	0.446	0%	34.3
	One effect per study (smallest)	9	0.53 (0.19-0.86)	0.002	77%	5.6
	One effect per study (largest)	9	0.61 (0.28-0.93)	<0.001	71%	4.8
	Trim-and-fill procedure	6	0.24 (-0.06 to 0.55)	-	-	13.5
	Social anxiety primary target/outcome	6	0.74 (0.24-1.24)	0.003	83%	3.9
	Pre-selected for social anxiety symptoms	5	0.75 (0.16-1.32)	0.011	86%	3.8
	CBT apps only	4	0.73 (0.01-1.45)	0.044	82%	3.9
Obsessive-compulsive symptoms						
	Apps vs. control conditions	5	0.51 (0.18-0.84)	0.002	42%	5.8
Panic symptoms						
	Apps vs. control conditions	2	-0.12 (-0.50 to 0.25)	0.515	0%	-
Acrophobia symptoms						
	Apps vs. control conditions	2	0.90 (0.38-1.42)	0.001	0%	3.0

n – number of comparisons, NNT – number needed to treat, CBT – cognitive behavioral therapy

stress symptoms was $g=0.12$ (95% CI: 0.03-0.21, $p=0.007$), corresponding to an NNT of 28.3. Heterogeneity was low ($I^2=24%$). Significant small effects were observed in all sensitivity analyses, except when smaller sample trials were removed.

The pooled effect size for the ten comparisons between apps ($N=576$) and control conditions ($N=447$) on social anxiety symptoms was $g=0.52$ (95% CI: 0.22-0.82, $p=0.001$), corresponding to an NNT of 5.7. Heterogeneity was high ($I^2=75%$). Effects remained stable and similar in magnitude when restricting the analyses to trials that pre-selected participants for social anxiety symptoms, delivered a CBT app, and where social anxiety was the primary target or outcome. Non-significant effects were observed when restricting the analyses to trials with a lower risk of bias rating and a larger sample. However, only one trial was rated as low risk and having a larger sample³⁷.

Significant pooled effect sizes were observed for the five comparisons between apps and control conditions on obsessive-compulsive symptoms ($g=0.51$, 95% CI: 0.18-0.84, $p=0.002$; NNT=5.8)

and for the two comparisons on acrophobia symptoms ($g=0.90$, 95% CI: 0.38-1.42, $p=0.001$; NNT=3.0). A non-significant negative effect size was observed for the two comparisons between apps and control conditions on panic symptoms ($g=-0.12$, 95% CI: -0.50 to 0.25, $p=0.515$) (see Table 4).

Dropout rates

From 182 conditions with available data, the weighted dropout rate was estimated to be 23.6% (95% CI: 21.3-26.1, $I^2=93%$). When removing small sample studies, the dropout rate was 29.9% (95% CI: 26.0-34.0, $I^2=96%$) from 67 conditions. When restricting the analyses to lower risk of bias studies, the dropout rate was 26.6% (95% CI: 22.7-31.0, $I^2=94%$) from 50 conditions. For samples of participants with depression, the dropout rate was 28.7% (95% CI: 23.6-34.3, $I^2=93%$) from 42 conditions. For samples of participants with anxiety, the dropout rate was 25.4% (95% CI: 20.6-31.0,

$I^2=92%$) from 39 conditions. For CBT apps, the dropout rate was 23.3% (95% CI: 19.8–27.2, $I^2=94%$) from 89 conditions.

DISCUSSION

Interest in mental health apps as a scalable tool to treat symptoms of depression and anxiety continues to grow. Since the last comprehensive meta-analysis published in 2019¹³, more than 100 RCTs have been conducted. To ensure that clinicians, policymakers and the public have access to the latest information on the evidence base of these apps, we conducted an updated meta-analysis of 176 research trials. A particular focus was on identifying features that may account for the evident and considerable heterogeneity in efficacy from study to study. This is the first meta-analysis of mental health apps to undertake a thorough analysis of how combinations of putative factors may interact, in order to provide new insights into the circumstances and subgroups of individuals for which certain app features may confer greatest effects.

Overall, results showed that mental health apps have overall small but significant effects on symptoms of depression ($N=33,576$, $g=0.28$) and generalized anxiety ($N=22,394$, $g=0.26$), corresponding to an NNT of 11.5 and 12.4, respectively. Heterogeneity was high in main analyses, but substantially lower when removing outliers. Effects were robust across extensive sensitivity analyses, and were similar in magnitude at different follow-up lengths and after removing small sample and higher risk of bias trials. Larger effects were found for depression when it was the primary target of the app ($g=0.38$), while this was not the case for generalized anxiety ($g=0.20$). Attrition was apparent, with one in four participants prematurely dropping out of their allocated app program. Small non-significant effects for depression and generalized anxiety were observed when evaluating apps against web and face-to-face interventions, though the number of studies was low and confidence intervals wide.

There was less variability in outcome scores at post-test in app compared to control conditions ($RoV=-0.14$ for depressive symptoms and $RoV=-0.21$ for generalized anxiety symptoms). However, heterogeneity was high for differences in variance ($I^2=78%$ for depressive symptoms and $I^2=75%$ for generalized anxiety symptoms), suggesting variable efficacy of app interventions.

The expanding literature now enables us to assess the effects of mental health apps on specific symptoms of anxiety, and highlights the potential of more specialized approaches. Our previous meta-analytic estimates of apps on symptoms of social anxiety, panic and post-traumatic stress were only based on three to six comparisons, finding limited evidence of efficacy¹³. Now the literature has evolved, with the number of trials targeting certain symptoms more than tripling (e.g., post-traumatic stress), while newer trials have emerged that enable calculation of preliminary pooled effects for other symptoms (e.g., obsessive-compulsive symptoms, acrophobia). We found evidence of moderate effects of apps on social anxiety ($n=10$, $g=0.52$) and obsessive-compulsive ($n=5$, $g=0.51$) symptoms, a small effect on post-traumatic stress

symptoms ($n=17$, $g=0.12$), a large effect on acrophobia symptoms ($n=2$, $g=0.90$), and a non-significant negative effect on panic symptoms ($n=2$, $g=-0.12$). However, these results should be considered with caution, because most trials contributing to these analyses had considerable risk of bias and were based on small sample sizes.

Our results also highlight how advances in trial methodology will aid in better assessing the efficacy of extant apps and the design of new ones. At the univariate level, the type of control condition emerged as a moderator, with inactive controls generating larger effects on depression and generalized anxiety (identified as primary target) than placebo controls or care as usual. This is a well-replicated finding observed across all modes of psychological treatment^{38–41}, and provides confirmation that some of the benefits of apps are explained by the “digital placebo” effect⁴². Now that this placebo effect has been well established, an imperative arises to ensure that its real-world implications are realized by both the clinicians who assess the efficacy or “prescribe” certain apps and the regulators who certify their claims. Furthermore, effects were larger among pre-selected samples of participants with threshold-level symptoms of depression and generalized anxiety (identified as primary target). This finding aligns with prior meta-analytic evidence that higher baseline severity is associated with better outcomes of digital interventions^{43,44}, suggesting that people with moderate to severe depression or anxiety at baseline benefit more from care augmented by these apps⁴⁵.

Features of the app were also associated with effect sizes at the univariate level. Apps that were based on CBT produced larger effects than other apps, such as those based on mindfulness. This finding is not surprising, given that the evidence base for all forms of CBT in depression and anxiety is substantially larger than for other approaches⁴⁶. Perhaps a better insight towards the “active ingredients” of CBT for these conditions (compared to other approaches) has facilitated the development of apps that prioritize these effective components over other less effective or potentially harmful ones.

Furthermore, we found some evidence that effects were larger when apps specifically designed for depression incorporated chatbot technology, and when apps specifically designed for anxiety incorporated mood monitoring features. It is possible that these innovative technological features offer a greater degree of personalization, are more engaging, foster emotional self-awareness, and keep users more accountable for making progress^{47,48}, potentially resulting in greater benefit. However, these results were derived from *post-hoc* analyses and should be considered with caution, because the number of studies in these subgroups was relatively small. Randomized experiments that test the added effects of chatbot and mood monitoring technology as both mechanisms of action and drivers of engagement are needed.

At a multivariate level, combinations of proposed moderator variables helped to better identify subgroups where apps were more or less efficacious. For depression, studies with higher efficacy estimates tended to be characterized by pre-selected samples of smaller size and without a formal psychiatric diagnosis. Cognitive training apps had higher efficacy for depression when paired with

these trial features ($g=1.16$), but we emphasize the need for caution in interpreting this effect, given the paucity of studies in this grouping ($n=3$). For generalized anxiety, apps with mood monitoring were particularly efficacious for studies with pre-selected samples, while the positive benefits of this monitoring did not emerge for more universal samples.

There are some limitations to this meta-analysis that should be considered. First, analyses were restricted to the post-intervention period, because of inconsistent reporting and length of follow-up assessments, and variability in how dropouts were dealt with. Thus, whether the benefits of apps observed in the short term extend over longer periods remains an open question. Second, analyses could only be conducted on symptom change and not on other, clinically meaningful outcomes such as remission, recovery, or deterioration⁴⁹. Despite a sufficient number of trials sampling individuals with depression or anxiety, very few reported these outcomes at all or, if they did, they defined them inconsistently. Third, heterogeneity was high in many of the main analyses. Even though we tried to explain this through subgroup analyses, we were not able to explain all of it, and many of the subgroups still had high heterogeneity. However, it is perhaps inevitable that some heterogeneity will always persist when aggregating data from trials of essentially different apps – regardless of how well the individual components and trial characteristics are categorized.

In conclusion, we present the most comprehensive meta-analysis of the effects of mental health apps on symptoms of depression and anxiety, including 176 trials. We conclude that apps have overall small but significant effects on symptoms of depression and generalized anxiety. Larger effects are observed in trials in which depression is the primary intervention target or outcome, suggesting that apps could be a suitable first step in treatment for those receptive to this approach or those who cannot access traditional forms of care. Certain features of apps, such as mood monitoring and chatbot technology, were associated with larger effect sizes, although this needs to be confirmed in future experimental research. Evidence supporting the efficacy of apps for specific anxiety symptoms is uncertain, largely due to trials with considerable risk of bias and small sample sizes. As responsiveness to mental health apps varies, future research would benefit from collecting and pooling large datasets (with passive and self-reported data) to generate predictive models capable of accurately detecting those for whom an app is sufficient from those who require different forms of treatment.

ACKNOWLEDGEMENTS

J. Linardon holds a National Health and Medical Research Council Grant (APP1196948); J. Torous is supported by the Argosy Foundation; J. Firth is supported by a UK Research and Innovation Future Leaders Fellowship (MR/T021780/1). Supplementary information on this study is available at <https://osf.io/ufb2w/>.

REFERENCES

- Santomauro DF, Herrera AMM, Shadid J et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 2021;398:1700-12.
- Liu Q, He H, Yang J et al. Changes in the global burden of depression from 1990 to 2017: findings from the Global Burden of Disease study. *J Psychiatr Res* 2020;126:134-40.
- Cuijpers P, Donker T, Weissman MM et al. Interpersonal psychotherapy for mental health problems: a comprehensive meta-analysis. *Am J Psychiatry* 2016;173:680-7.
- Cuijpers P, Miguel C, Harrer M et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry* 2023;22:105-15.
- Carl E, Witcraft SM, Kauffman BY et al. Psychological and pharmacological treatments for generalized anxiety disorder (GAD): a meta-analysis of randomized controlled trials. *Cogn Behav Ther* 2020;49:1-21.
- Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-66.
- Cuijpers P, Sijbrandij M, Koole SL et al. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry* 2013;12:137-48.
- Torous J, Bucci S, Bell IH et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 2021;20:318-35.
- Torous J, Myrick KJ, Rauseo-Ricupero N et al. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR Mental Health* 2020;7:e18848.
- Andersson G. Internet-delivered psychological treatments. *Annu Rev Clin Psychol* 2016;12:157-79.
- Hedman-Lagerlöf E, Carlbring P, Svärdman F et al. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *World Psychiatry* 2023;22:305-14.
- Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center* 2016;22:1-44.
- Linardon J, Cuijpers P, Carlbring P et al. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* 2019;18:325-36.
- Goldberg SB, Lam SU, Simonsson O et al. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLoS Digit Health* 2022;1:e0000002.
- Baumel A, Torous J, Edan S et al. There is a non-evidence-based app for that: a systematic review and mixed methods analysis of depression-and anxiety-related apps that incorporate unrecognized techniques. *J Affect Disord* 2020;273:410-21.
- Larsen ME, Huckvale K, Nicholas J et al. Using science to sell apps: evaluation of mental health app store quality claims. *NPJ Digit Med* 2019;2:18.
- Torous J, Myrick K, Aguilera A. The need for a new generation of digital mental health tools to support more accessible, effective and equitable care. *World Psychiatry* 2023;22:1-2.
- Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. *J Consult Clin Psychol* 2020;88:1-13.
- Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. *JMIR mHealth uHealth* 2015;3:e4924.
- Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. *JAMA Netw Open* 2022;5:e2248784.
- Firth J, Torous J, Nicholas J et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry* 2017;16:287-98.
- Cuijpers P, Griffin JW, Furukawa TA. The lack of statistical power of subgroup analyses in meta-analyses: a cautionary note. *Epidemiol Psychiatr Sci* 2021;30:e78.
- Firth J, Torous J, Nicholas J et al. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J Affect Disord* 2017;218:15-22.
- Mills HL, Higgins JPT, Morris RW et al. Detecting heterogeneity of intervention effects using analysis and meta-analysis of differences in variance between trial arms. *Epidemiology* 2021;32:846-54.
- Usui T, Macleod MR, McCann SK et al. Meta-analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research. *PLoS Biology* 2021;19:e3001009.
- Li X, Dusseldorp E, Meulman JJ. Meta-CART: a tool to identify interactions

- between moderators in meta-analysis. *Br J Math Stat* 2017;70:118-36.
27. Moher D, Liberati A, Tetzlaff J et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151:264-9.
 28. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley, 2011.
 29. Borenstein M, Hedges LV, Higgins JP et al. *Introduction to meta-analysis*. Chichester: Wiley, 2009.
 30. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. San Diego: Academic Press, 1985.
 31. Cohen J. A power primer. *Psychol Bull* 1992;112:155-9.
 32. Harrer M, Cuijpers P, Furukawa TA et al. *Doing meta-analysis with R: a hands-on guide*. New York: Chapman and Hall/CRC, 2021.
 33. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56: 455-63.
 34. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-4.
 35. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
 36. Li X, Dusseldorp E, Meulman JJ. A flexible approach to identify interaction effects between moderators in meta-analysis. *Res Synth Methods* 2019;10:134-52.
 37. Bruehlman-Senecal E, Hook CJ, Pfeifer JH et al. Smartphone app to address loneliness among college students: pilot randomized controlled trial. *JMIR Ment Health* 2020;7:e21496.
 38. Linardon J. Can acceptance, mindfulness, and self-compassion be learnt by smartphone apps? A systematic and meta-analytic review of randomized controlled trials. *Behav Ther* 2020;51:646-58.
 39. Michopoulos I, Furukawa TA, Noma H et al. Different control conditions can produce different effect estimates in psychotherapy trials for depression. *J Clin Epidemiol* 2021;132:59-70.
 40. Christ C, Schouten MJ, Blankers M et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. *J Med Internet Res* 2020;22:e17831.
 41. Linardon J, Shatte A, Messer M et al. E-mental health interventions for the treatment and prevention of eating disorders: an updated systematic review and meta-analysis. *J Consult Clin Psychol* 2020;88:994-1007.
 42. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. *Lancet Psychiatry* 2016;3:100-2.
 43. Bower P, Kontopantelis E, Sutton A et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ* 2013;346:f540.
 44. Scholten W, Seldenrijk A, Hoogendoorn A et al. Baseline severity as a moderator of the waiting list-controlled association of cognitive behavioral therapy with symptom change in social anxiety disorder: a systematic review and individual patient data meta-analysis. *JAMA Psychiatry* 2023;80:822-31.
 45. van Straten A, Hill J, Richards D et al. Stepped care treatment delivery for depression: a systematic review and meta-analysis. *Psychol Med* 2015;45:231-46.
 46. David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Front Psychiatry* 2018;9:4.
 47. Torous J, Lipschitz J, Ng M et al. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J Affect Disord* 2020;263:413-9.
 48. He Y, Yang L, Qian C et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res* 2023;25:e43862.
 49. Cuijpers P, Karyotaki E, Ciharova M et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. *Acta Psychiatr Scand* 2021;144:288-99.
- DOI:10.1002/wps.21183