

Review article

Association between user engagement and clinical outcomes in smartphone apps for depression and anxiety: A systematic review and meta-analysis

Jake Linardon^{a,*}, John Torous^b, Mariel Messer^a, Claudia Liu^a, Imogen H. Bell^{c,d}, Jennifer Nicholas^{c,d}, Simon B. Goldberg^e

^a SEED Lifespan Strategic Research Centre, School of Psychology, Faculty of Health, Deakin University, Geelong, VIC, Australia

^b Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

^c Orygen, Parkville, VIC, Australia

^d Centre for Youth Mental Health, University of Melbourne, Melbourne, VIC, Australia

^e Department of Counseling Psychology and Center for Healthy Minds, University of Wisconsin, Madison, WI, United States

ARTICLE INFO

Keywords:

Smartphones

Depression

Anxiety

Meta-analysis

Engagement

ABSTRACT

Apps targeting symptoms of depression and anxiety have a growing evidence base for their efficacy, yet it remains unclear whether increased user engagement is necessary to enhance their benefits. This systematic review and meta-analysis examined the current evidence on the association between engagement and clinical outcomes in trials of depression and anxiety apps. We included original or secondary analyses of randomized trials of apps delivered to participants with elevated depression or anxiety that reported engagement-outcome relationships. Studies were identified through a previous systematic review and an updated search. Twenty-eight studies met inclusion criteria for the systematic review, and 13 were included in the meta-analysis. Qualitative synthesis revealed heterogeneity in how engagement-outcome associations were reported: >40 different engagement metrics were identified, and 57 % of studies examined multiple metrics as predictors. Averaging across engagement and outcome variables, a significant pooled effect was found ($r = 0.16$; 95 % CI= 0.09, 0.21), indicating that greater engagement was linked to larger symptom improvement. This effect remained significant following publication bias adjustment and when assuming a zero effect for seven studies that reported non-significant associations but no accompanying data ($r = 0.11$, 95 % CI= 0.06, 0.16). Significant effects were also observed when modeling specific engagement metrics, symptom outcomes, and app characteristics, although few studies contributed to these analyses. Engagement may play a small role in symptom improvement within depression and anxiety apps. Findings highlight the need for better reporting standards, including which theory-driven engagement metrics should be routinely reported in future research.

1. Introduction

Depression and anxiety are leading causes of disability worldwide (Santomauro et al., 2021), yet most individuals affected go untreated (World Health Organization, 2022). Interventions translated for delivery via smartphone applications (apps) can increase access to evidence-based psychological treatments due to their low cost, widespread availability, and ability to be used discreetly (Torous et al., 2021, 2025). This discreet access can help reduce stigma-related barriers, which remain a major deterrent to seeking face-to-face care (Kazdin, 2017). Apps can also incorporate varying degrees of personalisation, ranging from simple surface-level tailoring (e.g., use of the individual's

name) to more substantive adaptations, such as recommending specific modules or exercises based on users' reported symptoms, preferences, or progress (Linardon and Torous, 2025). Their portability allows individuals to access therapeutic content and apply learned skills in real-world contexts and moments of need, supporting the integration of treatment strategies into daily life.

While evidence supports the efficacy of apps for depression and anxiety as both a stand-alone option or an adjunct to traditional care (Linardon, Torous, et al., 2024), sustained user engagement remains a key concern (Linardon and Fuller-Tyszkiewicz, 2020; Torous et al., 2018). Clinical trials show that up to 60 % of participants fail to complete all prescribed modules, and as many as 70 % disengage within a

* Corresponding author.

E-mail address: jake.linardon@deakin.edu.au (J. Linardon).

<https://doi.org/10.1016/j.psychres.2025.116864>

Received 15 August 2025; Received in revised form 16 October 2025; Accepted 24 November 2025

Available online 25 November 2025

0165-1781/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

few weeks of use (Peake et al., 2024; Sirbu et al., 2025). Engagement issues are even more pronounced in naturalistic settings, with seminal work indicating that 30-day retention rates for popular mental health apps can be as low as 3 %, and daily active user rates hover around just 4 % (Baumel et al., 2019).

These findings have prompted a new generation of research aimed at investigating theories of engagement (Nahum-Shani et al., 2022) and developing effective engagement strategies, with promising evidence found for personalised push notifications (Bidargaddi et al., 2018), gamification techniques (Looyestyn et al., 2017), digital navigators (Perret et al., 2023), and peer support components (Fortuna et al., 2019). However, it may be premature to invest heavily in testing or optimising engagement strategies without first establishing whether increased engagement is indeed necessary to maximise the clinical benefits of mental health apps. Previous reviews of engagement in mental health apps have primarily focused on its reporting, measurement, and validity (Ng et al., 2019; Torous et al., 2020, 2018), but have not yet systematically examined the extent to which engagement – or specific engagement metrics – are associated with clinical outcomes. On the other hand, prior reviews of web-based mental health interventions highlight the complexity of the assumed engagement-outcome relationship, suggesting that the strength of the association may depend on how engagement is measured (Donkin et al., 2011; Gan et al., 2021). For example, greater program completion and total website exposure appear to be consistently linked to improved outcomes for depression, whereas metrics such as number of logins, self-reported task completions, time spent online, and pages viewed show no consistent association (Donkin et al., 2011).

While earlier reviews addressing this question offer important insights into the potential impact of user engagement on mental health outcomes, their exclusive focus on web-based interventions limits the generalisability of findings to smartphone apps targeting depression and anxiety. Web-based programs are typically accessed via desktop computers and follow a structured, modular-based format that emphasises psychoeducation and sequential learning (Andersson, 2024). However, over the past five years, this distinction has become less pronounced, as many digital mental health programs are now designed with a mobile-first approach and can be accessed on smartphones through responsive websites or progressive web apps, which often replicate much of the app experience. The primary differences tend to relate to native functionality, such as push notifications, offline access, and integration with smartphone sensors, which can influence both engagement opportunities and how engagement is measured (Torous et al., 2019). For example, whereas web program trials have historically reported metrics such as module completion or time on site (Donkin et al., 2011), app trials often include additional metrics like number of opens, session duration, and activity frequency (Ng et al., 2019). Given these overlapping but not identical design and functionality characteristics, findings from earlier web-based programs cannot be assumed to fully generalise to modern app-based interventions, highlighting the need for research that specifically examines engagement-outcome associations in trials of mental health apps.

The purpose of this review is to systematically examine the association between engagement and clinical outcomes in randomized controlled trials (RCTs) of mental health apps targeting depression and anxiety. More specifically, we aim to (1) synthesise research reporting on the associations between engagement and outcomes; (2) characterize the types of engagement metrics assessed in relation to predicting outcomes; and (3) quantify the extent of the engagement-outcome association through meta-analytic techniques, and explore whether different metrics of engagement (e.g., days of use, number of activities completed, time spent on app, etc.) are more strongly related to symptom change.

2. Method

2.1. Identification and selection of studies

We pre-registered this review in the Open Science Framework repository (<https://osf.io/xbh96>) and adhered to the PRISMA guidelines (Page et al., 2021). Pre-registration was submitted after the search was conducted but prior to the full-text screening and data extraction process. We first identified potentially eligible studies from a recent 2024 review on adverse events in clinical trials of mental health apps (Linardon, Fuller-Tyszkiewicz, et al., 2024). This review captured all available randomized trials on depression and anxiety apps up to 2024 that would meet eligibility for this research (search terms in the supplementary materials). We then updated the search by searching the Medline and PsycINFO databases from January 2024 to March 2025 using the following combinations of key terms: smartphone* “mobile phone” “mobile app*” “iphone” “android” “mhealth” “m-health” “mobile device*” “mobile-based” “mobile health” “tablet-based” app-based app-supported app-assisted AND random* “clinical trial” AND anxiety, anxious, phobia,* panic agoraphobia “mental health” “mental illness*” “depress*” “affective disorder*” “mood disorder*” mood. A secondary search strategy was employed by screening the reference lists of included trials and relevant reviews in this area. Furthermore, because analyses on engagement-outcome associations are often conducted *post hoc* and may be reported in secondary publications rather than the original trial reports, we also reviewed all available records that cited each trial of depression/anxiety apps to ensure that no relevant companion studies were missed.

Eligibility criteria were defined using the PICO framework:

Population: Individuals with elevated depression and/or anxiety, established through a diagnostic interview, scoring above a cut-off on a validated self-report scale, or by participant self-report. In this review, “elevated” refers to baseline symptom levels that either exceeded established clinical or subclinical cut-off scores on validated measures or were self-reported by participants as clinically significant difficulties, indicating elevated symptomatology rather than change from a prior timepoint.

Intervention: Native mental health smartphone applications designed to address symptoms of depression and/or anxiety. Apps could be delivered as either stand-alone interventions or as adjuncts to traditional clinical services.

Comparator: Any comparator condition, including waitlist, usual care, attention control, or alternative app-based interventions.

Outcome: Clinical outcomes related to depression and/or anxiety and their association with objective app engagement metrics (e.g., number of logins, days of use, activities completed). Studies relying solely on retrospective self-reports of engagement were excluded. Two researchers performed the screening at the full-text stage to determine whether the paper met full inclusion criteria. Inter-rater agreement was excellent ($\kappa = 0.97$).

2.2. Risk of bias and data extraction

Risk of bias was assessed using five domains from the Cochrane Risk of Bias tool (Higgins et al., 2019): random sequence generation; allocation concealment; blinding of participants or personnel, blinding of outcomes; and completeness of outcome data. Each domain received either a high, low or unclear rating. We also extracted the following information from eligible studies: target sample; sample selection criteria; sample size; app name; app orientation (e.g., CBT vs non-CBT – an app was coded as CBT if it explicitly reported being based on CBT principles or if its central therapeutic component involved cognitive restructuring); presence of key in-app features (guidance, symptom monitoring technology, and chat-bot); treatment delivery mode (stand-alone or adjunctive); type of control group; engagement metrics

and their operationalization; outcome variables; follow-up length; engagement-outcome relationship data; and a summary of findings. Two researchers performed data extraction, with minor discrepancies ($\kappa > 0.82$) resolved through consensus.

2.3. Analytical approach

We first performed a narrative synthesis of findings regarding associations between engagement metrics and clinical outcomes. Findings were synthesized overall and then by target symptom outcome. We focused on the proportion of significant and non-significant relationships identified in this qualitative synthesis.

Meta-analyses were then performed to quantify associations between engagement and symptom improvement within the app arm among those trials that reported sufficient data to calculate effect sizes ($k = 13$). Pearson's correlation coefficient (r) was selected as the measure of effect size, with values of 0.1 considered small, 0.3 considered medium, and 0.5 considered large (Cohen, 1992). One trial provided beta weights (Mohr et al., 2019), which were then converted to r based on prior recommendations (Peterson and Brown, 2005). For the two trials that dichotomized engagement (high vs. low; Greer et al., 2019; Mantani et al., 2017), standardized mean differences were calculated first and then converted to r through standard methods implemented in the Comprehensive Meta-Analysis (CMA) software. Correlation coefficients were transformed prior to analyses using Fisher's Zr-transformation so that each effect size could be weighted by its inverse variance (Lipsey and Wilson, 2001). These effect sizes were converted back into standard correlation coefficients when reporting results. All correlations were standardized such that a positive coefficient indicated that greater engagement was associated with larger symptom improvement.

A total pooled effect for the association between engagement and symptom change was computed by first aggregating within-study correlations across engagement metrics and outcome measures at post-test. This within-study averaging was conducted using CMA, which applies a sample size-weighted approach before pooling effects in the overall meta-analysis. Engagement metrics included in the meta-analyses encompassed number of days of use, number of activities/tasks completed, number of app sessions, and time spent on the app, as these were consistently reported among the 13 trials eligible for analysis. We then conducted a series of sensitivity analyses, calculating pooled effects separately (where feasible) for those specific engagement metrics and for specific symptom outcomes.

Random effects models were used. Heterogeneity was assessed through 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 variance (Higgins and Thompson, 2002). We employed the trim-and-fill procedure (Duval and Tweedie, 2000) to evaluate the impact of potential publication bias. We considered additional publication bias tests (e.g., precision-effect test and precision-effect estimate with standard errors [PET-PEESE]), but did not pursue this due to poor performance of these methods in the context of small numbers of studies (e.g., $k < 20$), especially when the true effect is small (Stanley, 2017). However, in a conservative sensitivity analysis, we assumed an effect size of $r = 0.00$ for studies that reported finding a non-significant association between engagement and outcomes but did not report an exact effect size or other usable data (e.g., the paper merely reported that associations were n.s or that $p > .05$). Table 1 outlines these studies.

Pre-registered subgroup analyses were not conducted given the limited number of studies available for meta-analysis, but sensitivity analyses were performed by computing effects for specific trial features to see if the total effects were robust under various conditions.

3. Results

A flowchart of the literature search is presented in Fig. 1. Twenty-

four studies from the prior review met the full inclusion criteria, and an additional four studies were identified through the updated search, bringing the total to 28 studies included in the systematic review. Four of these trials conducted secondary analyses on dose-response associations reported in companion publications. Thirteen studies provided sufficient to calculate effect sizes for meta-analysis (Table 1). Seven studies reported a non-significant engagement-outcome association but did not provide data necessary to calculate effects (Araya et al., 2021; Catuara-Solarz et al., 2022; Miklowitz et al., 2023; Newman et al., 2021; Oh et al., 2020; Stiles-Shields et al., 2018; Stiles-Shields et al., 2024; Tighe et al., 2017), so these were included in the sensitivity meta-analyses that assumed an $r = 0.00$.

3.1. Study characteristics

Table 1 presents the characteristics of the included studies. Samples comprised participants with elevated depression ($k = 13$), generalized anxiety ($k = 4$), mixed anxiety and depression¹ ($k = 5$), social anxiety ($k = 3$), panic ($k = 1$), and specific phobic ($k = 2$) symptoms. Most trials screened participants based on self-report ($k = 18$) rather than diagnostic interviews ($k = 10$). There were 34 app conditions in total, most of which were based on CBT ($k = 27$), had symptom tracking features ($k = 21$), and were delivered in guided self-help format ($k = 19$). Few trials incorporated chatbot technology ($k = 2$). More trials delivered the app as a stand-alone intervention ($k = 22$) rather than an adjunct to more intensive treatment ($k = 6$). Waitlists were the most common type of control conditions ($k = 12$). The total sample size randomized ranged from 30 to 1312. Risk of bias ratings varied across studies. Twenty-one trials met criteria for adequate sequence generation, 11 met criteria for adequate allocation concealment, eight met criteria for proper participant blinding (e.g., in a way ensuring that participants were not aware of the condition they were allocated to), 27 used blinding outcome assessors or self-report outcome measures, and 18 met criteria for appropriate handling of missing data. Two trials were identified as low risk of bias across all criteria, eight met four, 11 met three, three met two, and four trials met one of the criteria.

3.2. Narrative synthesis

3.2.1. All studies

There was variability in the number of engagement metrics tested for its association with outcomes. The minimum and maximum number of metrics reported was one and 10, with most trials ($k = 13$) only reporting one metric. The mean number of metrics reported from 28 studies was 2.3 (SD = 2.0).

Table 2 presents the number of statistically significant and non-significant associations between different engagement metrics and symptom outcomes across included studies. Engagement metrics were divided into five broad categories: activity/task completion; days of use; time of use; number of sessions/opens; and "other". Across all samples, most studies used at least one measure of activity/task completion ($k = 18$), with 34 associations with symptom improvement tested. Of these, eight were statistically significant and in the expected direction. Number of app opens/sessions was used in nine studies, with 13 associations tested, three of which were significant and in the expected direction. Days of app use and time of use were, respectively, used in six and seven studies, with seven associations tested each. Of these, only one association was statistically significant for both metrics. "Other" engagement metrics included, for example, time to last use, module completions, and click rate, of which there were three of 12 associations that were statistically significant. Below, we provide a more detailed narrative

¹ Mixed anxiety and depression refer to samples comprising individuals with elevated symptoms of depression, anxiety, or both, even if these conditions did not necessarily co-occur within the same individual.

Table 1
Characteristics of included studies.

Author	Sample	App features (n)	Control (n)	RoB	Outcome variable	Engagement metric(s)
Araya et al. (2021) – Trial 1 **	Depression (PHQ-9 ≥ 10)	CONEMO (657) CBT Guided Symptom monitoring: NR Chat-bot: NR Format: Stand-alone	Usual care (655)	+ + - SR +	Depression (50 % decrease in PHQ-9)	Activities completed (3 levels) 1. Zero (ns) 2. 1–10 (ns) 3. 11–27 (ns)
Araya et al. (2021) – Trial 2	Depression (PHQ-9 ≥ 10)	CONEMO (657) CBT Guided Symptom monitoring: NR Chat-bot: NR Format: Stand-alone	Usual care (655)	+ + - SR +	Depression (50 % decrease in PHQ-9)	Activities completed (3 levels) 1. Zero (+) 2. 1–10 (ns) 3. 11–27 (ns)
Bastiaansen et al. (2022)	Depression (interview)	Do-Module (55) Non-CBT Guided Symptom mon: yes Chat-bot: NR Format: adjunct Think Module (55) Non-CBT Guided Symptom mon: yes Chat-bot: NR Format: adjunct	Usual care (51)	? ? – SR +	Depression (IDS-SR)	Activities completed (2 levels) (+) 1. Compliant (≥75 % completion rate) 2. Not compliant (< 75 % completion rate)
Bell et al. (2023)*	Depression & anxiety (PHQ-8 ≥ 10 and GAD-7 ≥ 10)	Mello (29) CBT Unguided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (26)	+ + - SR +	Depression (PHQ-8) General anxiety (GAD-7)	Activities commenced (ns for dep & anx)* Check-ins completed (ns for dep & anx)* App use days (ns for dep & anx)* Proportion of app use days (ns for dep & anx)
Catuara-Solarz et al. (2022)**	Anxiety (GAD-7 score 5–18)	Foundations (95) CBT Unguided Symptom mon: NR Chat-bot: NR Format: stand-alone	Waitlist (95)	+ + - SR –	General anxiety (GAD-7)	App use days (ns) Time of app use (ns) Activities completed (ns)
Donker et al. (2019) reported in Donker et al. (2020)	Specific phobia (AQ ≥ 45.45)	Zerophobia (96) CBT Unguided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (97)	+ + - SR +	Acrophobia symptoms (AQ)	Time of app use (+) Activities completed (ns)
Graham et al. (2020)**	Depression & anxiety (PHQ-8 ≥ 10 and GAD-7 ≥ 8)	IntelliCare (74) CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (72)	+ ? – + +	Depression (PHQ-8) General anxiety (GAD-7)	Number sessions completed* (ns for dep & anx) Time to last use (ns for dep & anx) App use days* (ns for dep & anx)
Greer et al. (2019)	Anxiety (HADS > 7)	Not names (72) CBT Unguided Symptom mon: NR Chat-bot: NR Format: stand-alone	Placebo (73)	+ ? + ++	General anxiety (HAM-A) Depression (HADS-D)	Number sessions completed (ns anx & + dep) 1. Compliant (≥6 sessions) 2. Not compliant (> 6 sessions)
Guo et al. (2020) reported in Li et al. (2022)	Depression (CES-D ≥ 16)	Run4Love (150) CBT Unguided Symptom: NR Chat-bot: NR Format: stand-alone	Information resources (40)	+ ? – SR +	Depression (CES-D)	Number activities completed (2 levels) (+) 1. Compliant (average 74 % completed activities) 2. Not compliant (average of 15 % completed activities)
Kulikov et al. (2023)*	Depression (self-identified)	Spark Direct (35) CBT Unguided Symptom mon: yes Chat-bot: yes Format: adjunct	Placebo (25)	+ ? + SR –	Depression (PHQ-8)	Number activities completed (+) *
Lacey et al. (2023)*	Specific phobia (BSSSP ≥ 4)	oVRcome (63) CBT Unguided Symptom mon: NR	Waitlist (63)	+ ? – SR –	Phobic symptoms (SMSP)	Hours of app use (ns)*

(continued on next page)

Table 1 (continued)

Author	Sample	App features (n)	Control (n)	RoB	Outcome variable	Engagement metric(s)
Mantani et al. (2017) reported in Furukawa et al. (2018)*	Depression (interview)	Chat-bot: NR Format: stand-alone Kokora (81) CBT Guided Symptom mon: yes Chat-bot: NR Format: adjunct	Antidepressants (83)	+++ SR +	Depression (≥ 4 points change on PHQ-9)	Number sessions completed (ns)* Days to complete one session (ns) Time spent (ns)* Number mind map activities (ns)* Number behavioral activation activities (+)* Number cognitive restructuring activities (+)*
Miklowitz et al. (2023)**	Mood disorder (interview)	My Coach Connect Not CBT Guided Symptom mon: yes Chat-bot: NR Format: adjunct	Placebo (32)	+ ? + + ?	Depression/mania (PSRS)	Number of days with completed activities (ns) Number of weeks with completed activities (ns)
Moberg et al. (2019)*	Depression or anxiety (PHQ-9 or GAD-7 score 5–14)	Pacifica (253) CBT Unguided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (247)	? ? – SR +	Depression (PHQ-9) General anxiety (GAD-7)	Number of logins (ns dep & anx)* Number thought records complete (ns dep & anx)* Number goals complete (ns dep & anx) Number relaxations complete (ns dep & anx) Number community activities complete (ns dep & anx)
Mohr et al. (2019)*	Depression or anxiety (PHQ-9 ≥ 10 or GAD-7 ≥ 8)	IntelliCare (74, 76, 75, 76) CBT Guided/guided/unguided/unguided Symptom mon: NR Chat-bot: NR Format: stand-alone	-	? ? + SR +	Depression (PHQ-9) General anxiety (GAD-7)	Number app sessions (+ dep, ns anx)* Time to last app use (+ dep, ns anx) Number app downloads (+ dep & anx)*
Newman et al. (2021)**	Anxiety (GAD-Q criteria)	Not named (50) CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (50)	+ + - SR +	General anxiety (DASS-A, PSWQ & STAI-T composite)	Number app sessions (ns) Number messengers sent (ns) Number messengers received (ns) Number app visits (ns) Time spent on app (ns)
Oh et al. (2020)**	Panic disorder (interview)	Todaki (23) CBT Unguided Symptom: yes Chat-bot: yes Format: stand-alone	Information resources (22)	? ? – SR –	Panic symptoms (PDSS)	Time spent on app (ns) Days of app use (ns)
Raeuuri et al. (2021)*	Depression (interview)	Meru Health Program (63) CBT Guided Symptom mon: NR Chat-bot: NR Format: Adjunct	Usual care (61)	+ + + +	Depression (PHQ-8) Generalized anxiety (GAD-7)	Minutes of mindfulness practice (ns for dep & anx)* Number of messages sent to therapist (ns for dep & anx)
Roy et al. (2021)	Anxiety (interview)	Unwinding Anxiety (32) Non-CBT Guided Symptom mon: yes Chat-bot: NR Format: Adjunct	Usual care (33)	+ + - SR –	Generalized anxiety (GAD-7)	Number modules completed (+)
Saulnier et al. (2023)*	Social anxiety (ASI ≥ 6)	Boast (19) CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (17)	? ? – SR –	Social anxiety (ASI-3)	Number activities completed (+)* Number of completed EMA prompts (ns)
Schwob and Newman (2023)	Social anxiety (interview)	ImExposure (43) CBT Unguided Symptom mon: yes Chat-bot: NR Format: stand-alone	Placebo (39)	+ ? + SR +	Social anxiety (Composite SIAS & SPDQ)	Number activities completed (2 levels) (+)* 1. Compliant (≥ 3 times daily) 2. Not compliant (< 3 times daily)
Six et al. (2022)*	Depression (PHQ-8 ≥ 5)	AirHeart (45, 49) CBT Unguided/unguided Symptom mon: yes/yes	-	+ ? + SR ?	Depression (PHQ-9)	Number activities (journals) completed (ns)*

(continued on next page)

Table 1 (continued)

Author	Sample	App features (n)	Control (n)	RoB	Outcome variable	Engagement metric(s)
Stiles-Shields et al. (2018)**	Depression (PHQ-9 ≥ 10 & QIDS ≥ 11)	Chat-bot: NR Format: stand-alone Boost Me (10) CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone Thought Challenger (1) CBT Guided Symptom mon: NR Chat-bot: NR Format: stand-alone	Waitlist (10)	+ + - SR -	Depression (PHQ-9)	Number of app logins (ns) Number activities completed (ns)
Colleen Stiles-Shields et al. (2024)**	Depression or anxiety (PHQ-9 ≥ 10 or GAD-7 ≥ 8)	Vira (65) Not CBT Guided Symptom mon: yes Chat-bot: NR Format: Stand-alone	Placebo (65)	? ? - + +	Depression (PHQ-9) General anxiety (GAD-7)	Number of app logins (ns for dep & anx) Number of coaching interactions (ns for dep & anx)
Stolz et al. (2018)*	Social anxiety (interview)	Not named (60) CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (30) Web treatment (60)	+ ? - + +	Social anxiety (SPS, SIAS & LSAS composite)	Number of lessons completed (+) Number of hours spent in app (ns) * Number of separate sessions (ns) * Number of app clicks (ns) Number of relaxation exercises (ns) * Number of fear-provoking exercises (ns) * Number recorded helpful thoughts (ns) * Number exposure exercises (+) * Number mails written by patients (ns) Number of characters written by patients in mails (ns) Time spent on app (ns)
Tighe et al. (2017) reported in Tighe et al. (2020)**	Depression (PHQ-9 ≥ 10 or K-10 ≥ 25)	iBobby (31) Not CBT Guided Symptom mon: yes Chat-bot: NR Format: stand-alone	Waitlist (10)	+ ? - SR +	Depression (PHQ-9)	
Yatziv et al. (2024)	Depression (interview)	MoodVille (87) Not CBT Unguided Symptom: NR Chat-bot: NR Format: stand-alone	Waitlist (30)	? ? - SR -	Depression (MADRS)	Days of app use (2 levels) (+) 1. Compliant (≥ 4 days per week) 2. Not compliant (< 4 days per week)
Zuccolo et al. (2024)	Depression (interview)	Motherly (132) CBT Unguided Symptom: Yes Chat-bot: NR Format: stand-alone	Placebo (132)	+ + + SR +	Depression (EPDS)	Number activities completed (ns)

* indicates that the study/engagement metric was included in the meta-analysis. NR = not reported; PHQ-9 = Patient Health Questionnaire; EPDS = Edinburgh Postnatal Depression Scale; MADRS =Montgomery–Åsberg Depression Rating Scale; GAD-7 = Generalized Anxiety Disorder Scale; SPS = Social Phobia Scale; SIAS = Social Interaction Anxiety Scale; LSAS = Liebowitz Social Anxiety Scale; SPDQ = Social Phobia Diagnostic Questionnaire; PDSS= Panic Disorder Severity Scale; STAI-Straight-Trait Anxiety Scale; DASS = Depression Anxiety Stress Scale; AQ = Acrophobia Questionnaire; HADS = Hospital Anxiety Depression Scale; CES-D = Center for Epidemiologic Studies Depression Scale; IDS-SR = Inventory of Depressive Symptomatology – Self-Report.

Ns = non-significant; + = greater engagement associated with greater symptom change.

For RoB, ? = unclear, - = high risk, + = low risk. In order, sequence generation, allocation concealment, blinding of participants; use of self-report; and appropriate handling of missing data.

** indicates studies for which did not provide data to calculate effect sizes, but were coded as $r = 0$ in the sensitivity analyses.

synthesis by target symptom.

3.2.2. Depression

Nineteen studies used a measure of depressive symptoms as an outcome; of these, thirteen sampled participants with elevated depression (Araya et al., 2021; Bastiaansen et al., 2022; Guo et al., 2020; Kulikov et al., 2023; Mantani et al., 2017; Miklowitz et al., 2023; Rae-vuori et al., 2021; Six et al., 2022; Stiles-Shields et al., 2018; Tighe et al.,

2017; Yatziv et al., 2024; Zuccolo et al., 2024), five sampled a trans-diagnostic population with mixed depression and/or anxiety (Bell et al., 2023; Graham et al., 2020; Moberg et al., 2019; Mohr et al., 2019; Colleen Stiles-Shields et al., 2024), and one sampled participants with elevated anxiety but included depression as a secondary outcome (Greer et al., 2019). Table 2 shows that there were 42 associations tested between engagement and changes in depression, most of which comprised the operationalization of in-app activity/task completions ($n = 23$).

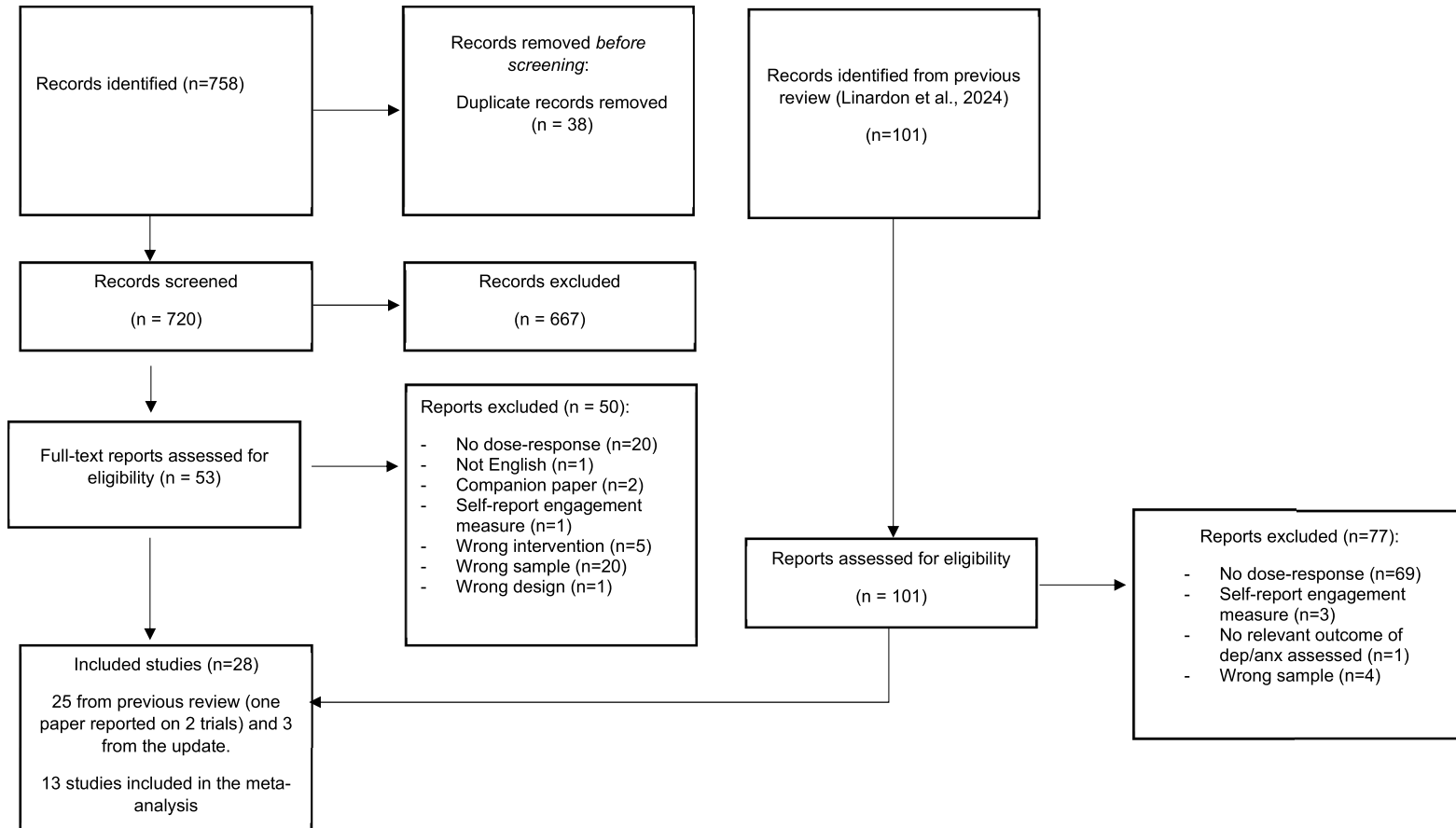


Fig. 1. Flowchart of literature search.

Table 2
Summary of Engagement Metric–Symptom Outcome Associations Across Studies.

Engagement Metric	Depression Sig/non-sig	General anxiety Sig/non-sig	Social anxiety Sig/non-sig	Specific phobia Sig/non-sig	Panic Sig/non-sig	Total Sig/non-sig
In-app activity/tasks completions	5/18	0/10	3/5	0/1	-	8/34
Days of app use	1/3	0/3	-	-	0/1	1/7
Time of app use	0/2	0/2	0/1	1/1	0/1	1/7
Number of sessions/app opens	2/5	1/7	0/1	-	-	3/13
Other	1/5	1/5	1/2	-	-	3/12

Significant relationships were all in the expected direction (higher engagement linked with greater improvement).

In-app activities comprised the following operationalisations: number check-ins, number of mind maps, behavioral activation, and cognitive restructuring activities, journal frequency, number of thought records, number of goals completed, number of relaxation exercises, number of community activities, number of messages sent to coach, total minutes of mindfulness practice, number of EMA prompts completed.

Time of app use comprised the following operationalisations: time per day average; total time of use; total hours of use; time spent on a session.

Number of sessions/app opens comprised the following operationalisations: number of app sessions, number of logins, number of downloads, number of visits, number of opens.

“other” comprised: proportion of days using the app; time to last use, days to complete a session, number of weeks with logged activity practice, number of messaged received from coach; number modules completed; coaching interaction frequency; number of clicks; number of characters written by patients.

Across all engagement metrics studied, nine associations were statistically significant and in the expected direction, including five related to in-app task completions. Two of five significant associations were in relation to number of app sessions. Thirty-three associations were statistically non-significant. No studies reported a significant association in a negative direction.

3.2.3. Generalized anxiety

Ten studies used a measure of generalized anxiety symptoms as an outcome; four sampled participants with elevated general anxiety (Catuara-Solarz et al., 2022; Greer et al., 2019; Newman et al., 2021; Roy et al., 2021), five sampled a mixed depression and/or anxiety sample (Bell et al., 2023; Graham et al., 2020; Moberg et al., 2019; Mohr et al., 2019; Stiles-Shields et al., 2018), and one reported generalized anxiety as a secondary outcome in a sample with depression (Raevuori et al., 2021). There were 29 associations tested between engagement and changes in generalized anxiety. Most of these operationalized engagement in terms of number of in-app activities ($n = 10$) or sessions ($n = 8$). Across all metrics, only two significant (and in the expected direction) associations were observed – one when engagement was operationalized as the number of downloads and the other when it was operationalized as the number of modules completed.

3.2.4. Social anxiety

Three studies used a measure of social anxiety symptoms as an outcome (Saulnier et al., 2023; Schwob and Newman, 2023; Stolz et al., 2018), all in samples screened for elevated social anxiety. Thirteen engagement-outcome associations were tested, eight of which operationalized engagement as number of in-app activities. Four associations were statistically significant and in the expected direction: three were in relation to number of in-app activities and the fourth was in relation to number of app lessons completed. The other nine associations were non-significant.

3.2.5. Specific phobia

Two trials sampled participants with a specific phobia (Donker et al., 2019; Lacey et al., 2023). One trial operationalized engagement in terms of time of app use and number of activities completed (Donker et al., 2019), while the other operationalized engagement as hours of app use (Lacey et al., 2023). The only significant association identified was for time of app use, with longer use predicting greater reduction in phobic symptoms.

3.2.6. Panic

One trial sampled participants with panic disorder (Oh et al., 2020). No significant associations were found between time spent on the app

and total days of use with changes in panic symptoms.

3.3. Meta-analysis

Table 2 presents the results from the meta-analyses. A small but significant total pooled effect size was observed from 13 studies ($r = 0.16$; 95 % CI = 0.09, 0.21), indicating that greater engagement was associated with larger symptom change. Heterogeneity was absent ($I^2 =$

Table 3
Results from the meta-analysis.

Analysis	<i>k</i>	<i>r</i> (95 % CI)	<i>p</i>	<i>I</i> ²
Total pooled effect	13	.16 (0.09, 0.21)	<0.001	0 %
Assumed <i>r</i> = 0 for non-sig. unreported studies	21	.11 (0.06, 0.16)	<0.001	0 %
Trim-and-Fill estimate	2	.14 (0.07, 0.21)		
Outliers removed	13 ^{trimmed}	.16 (0.09, 0.21)	<0.001	0 %
High risk of bias trials removed	12	.15 (0.09, 0.21)	<0.001	0 %
Engagement metrics				
Number of days of use	2	.11 (−0.15, 0.37)	.395	37 %
Number of completed tasks/activities	9	.19 (0.11, 0.27)	<0.001	0 %
Number of app sessions	6	.12 (0.05, 0.20)	.001	0 %
Time spent on the app	3	.17 (0.01, 0.31)	.027	0 %
Symptom outcomes				
Depression	9	.14 (0.07, 0.21)	<0.001	0 %
Generalized anxiety	6	.09 (0.01, 0.18)	.017	0 %
Social anxiety	4	.25 (0.08, 0.40)	.003	31 %
App features				
CBT app	12	.16 (0.09, 0.22)	<0.001	5 %
Guided app	5	.18 (0.02, 0.33)	.026	33 %
Unguided app	8	.15 (0.08, 0.22)	<0.001	0 %
Symptom monitoring features	8	.17 (0.07, 0.26)	.001	22 %
Stand-alone	10	.15 (0.08, 0.22)	<0.001	77 %
Augmented treatment	3	.19 (0.04, 0.34)	.012	0 %

0 %). There were no outliers (Table 3).

The impact of publication bias was examined using the trim-and-fill procedure. The pooled effect size remained significant and similar in magnitude when applying the trim-and-fill procedure ($r = .14$; 95 % CI = 0.07, 0.21). However, in a conservative sensitivity analysis that assumed non-significant effects were $r = 0.00$ for seven studies that reported a non-significant relationship without accompanying effect size data, the pooled effect was smaller but still significant ($r = 0.11$, 95 % CI = 0.06, 0.16).

Sensitivity analyses show that the effect remained robust when restricting the analyses to specific engagement and symptom outcome variables and when removing one trial with a high risk of bias rating. Significant effects were observed when engagement was operationalized as the number of activities completed ($k = 9$; $r = 0.19$; 95 % CI = 0.11, 0.27), number of sessions completed ($k = 6$; $r = 0.12$; 95 % CI = 0.05, 0.20), and time on the app ($k = 3$; $r = 0.17$; 95 % CI = 0.01, 0.31), but not from two studies that assessed engagement as days of use ($r = 0.11$; 95 % CI = -0.15, 0.37). Significant effects were also observed when changes in depression ($k = 9$; $r = 0.14$; 95 % CI = 0.07, 0.21), generalized anxiety ($k = 6$; $r = 0.09$; 95 % CI = 0.01, 0.18), and social anxiety ($k = 4$; $r = 0.25$; 95 % CI = 0.08, 0.40) were used as outcomes.

Effects also remained stable when limiting the analyses to specific app features. Significant, positive associations between engagement and symptom change were found for guided and unguided apps, apps with symptom monitoring technology, and apps delivered as either a stand-alone intervention option or an augmentation to traditional care ($r_s = 0.15$ to 0.19).

4. Discussion

This review of 28 studies synthesized empirical research investigating associations between engagement and clinical outcomes in randomized trials of mental health apps targeting depression or anxiety. Findings from our narrative synthesis highlight the complexities of the engagement-outcome association. Certain engagement metrics, particularly the number of activities or tasks completed, were more consistently linked with symptom improvement, especially for depression and social anxiety. Nonetheless, the heterogeneity in engagement operationalization, number of metrics assessed, and variability in sample size and study quality limited the ability to draw definitive conclusions from our qualitative synthesis regarding the role of engagement in symptom improvement.

Finding in-app activity completions to be one of the more consistent outcome predictors contrasts with Donkin et al.'s (2011) review of web-based interventions, which found no reliable association between these metrics and mental health symptom improvement. Importantly, their review relied on self-reported (not objective) activity completion, limiting the reliability of its conclusions. This is because while web-based platforms often capture objective metrics such as logins and module completion, they are limited in tracking skill use in daily life (Andersson, 2016). Unlike smartphone apps, they lack the portability and sensor integration needed to monitor behaviour in real-world settings. In contrast, apps routinely capture detailed user data in situ, allowing for more accurate engagement monitoring. As self-reported engagement – as synthesised by Donkin et al. (2011) – is an unreliable indicator of actual usage behaviour (Dwyer et al., 2024; Flett et al., 2019), this may account for the weaker associations observed in web-based trials. Second, the nature of the interventions themselves likely contributes to this divergence. Web programs are generally designed for prolonged, structured sessions involving substantial text and psychoeducation, making other metrics like overall site exposure a more relevant outcome predictor (Donkin et al., 2011). In contrast, apps are built for brief, flexible interactions that allow users to practice specific therapeutic skills in real time, making activity completion a more meaningful and proximal indicator of engagement and clinical benefit (Mohr et al., 2019).

Results from our meta-analysis provide preliminary support for the hypothesis that higher engagement is linked to greater clinical benefit. When averaging across engagement and outcome variables, we observed a small but significant pooled effect size of $r = 0.16$ (95 % CI = 0.09, 0.21) with no heterogeneity ($I^2 = 0$ %). This effect remained significant in various sensitivity analyses that adjusted for small-study bias, reporting bias, outliers, and risk of bias. Furthermore, significant pooled effects were observed when modelling different metrics of engagement (activity completions, number of sessions, time on app), symptom outcomes (depression, general anxiety, social anxiety), and in sensitivity analyses examining specific app characteristics, including CBT-only apps, guided and unguided formats, apps with symptom-tracking technology, and both stand-alone and adjunctive app interventions. While these results suggest that a potential dose-response relationship may apply across various contexts in mental health app trials, they should be interpreted with considerable caution, as many of the sensitivity analyses were based on a limited number of studies.

Broader clinical and research implications emerge from the current findings. The significant meta-analytic association observed suggests that user engagement to mental health apps might be an important therapeutic change mechanism, supporting continued efforts to develop and trial novel engagement strategies (Linardon and Fuller-Tyszkiewicz, 2020; Liu et al., 2025). Although small in magnitude, an effect of this size ($r = 0.16$) may still hold clinical relevance at the population level given the scalability and reach of app-based interventions; however, it also highlights that engagement alone is unlikely to drive meaningful clinical change in isolation. Several methodological factors may also help explain the modest strength of the association. These include restricted variability and skewed distributions in engagement metrics, the use of metrics (e.g., logins, session counts) that may not fully capture therapeutic engagement, the post hoc nature of most engagement-outcome analyses, and variability in intervention duration and timing of measurement. Other therapeutic factors such as the quality rather than quantity of engagement, degree of content personalization, digital working alliance, sudden shifts in emotional or cognitive processes (e.g., increased self-awareness or insight), user motivation, external support systems, or human relationships facilitated by apps may also prove to be important change mechanisms (Domhardt et al., 2025; Goldberg et al., 2022; Henson et al., 2019; Nahum-Shani et al., 2022). Maximising the clinical benefits of mental health apps will require future research to clarify the potential mechanisms of action, allowing for the purposeful design or refinement of interventions that prioritise the most effective therapeutic elements.

From a research perspective, the variability and suboptimal reporting of engagement-outcome associations in existing trials highlight the urgent need for improved reporting standards in this field. We noted that these associations are often examined *post hoc*, using multiple engagement metrics that are readily available but lack theoretical justification. This exploratory approach likely contributes to the inconsistent and limited reporting of relevant data in publications, which in turn hampers the ability to synthesise findings through meta-analytic techniques. Given the lack of progress in advancing standard engagement metrics for app research (Bradway et al., 2020; Nwosu et al., 2022), we recommend that the field work toward a consensus on which engagement metrics are most conceptually and clinically meaningful in the context of app-based interventions. Standardising these metrics will facilitate more consistent analyses across trials and enable the pooling of data in future meta-analyses to produce more precise and reliable estimates.

As an initial step, we suggest prioritising the reporting of number of activity completions and app logins, as these are currently the most commonly reported metrics, align well with the intended use of mental health apps (brief, repeated interactions throughout daily life), and are consistent with other reviews (Elkes et al., 2024; Forbes et al., 2023). Days of use may also prove to be a useful metric, given evidence that this variable may be less skewed than other usage metrics and may capture sustained engagement (Goldberg et al., 2025). We suggest that at a

minimum, all studies testing mental health apps report the correlation between the number of activities completed, app sessions, logins, and days of use with pre-post and pre-follow-up change in the study's primary mental health outcome. For instances where engagement metrics are highly skewed (e.g., >2; Curran et al., 1996), it may be valuable to dichotomize the sample into high and low usage categories and to report between-group comparisons on change in the primary outcome. These between-group comparisons could be converted into correlation coefficients in future meta-analyses. Furthermore, reporting all relevant statistical tests in-text or in supplementary materials, or making engagement data available via open science repositories, will enhance transparency and reproducibility, ultimately strengthening the evidence base (Nosek et al., 2015). It will also be important that authors do not selectively report only those associations that are significant. Preregistering hypotheses related to associations between engagement and symptoms may help guard against potential publication bias.

If data are also available to researchers, it would be useful to report engagement metrics that are more proximal to the therapeutic process. Examples include time spent engaging with core skill-building modules, completion of evidence-based therapeutic exercises (e.g., cognitive restructuring, behavioural activation tasks), frequency of mood tracking completions, and active use of in-the-moment coping or emotion regulation tools. These indicators may more accurately capture the quality and therapeutic relevance of app use, providing a stronger signal of clinically meaningful engagement than broader metrics like logins or downloads. Including such process-oriented metrics alongside standard usage indicators would help clarify which types of engagement are most strongly linked to treatment response and could guide the development of more targeted engagement strategies in future interventions. Equally important, reporting these associations regardless of whether they reach statistical significance is critical for building a cumulative evidence base. Non-significant findings provide essential information that can refine theory, reduce publication bias, and strengthen future meta-analytic estimates of engagement-response relationships.

The current findings must be interpreted within the context of its limitations. First, the number of studies available for analysis was low, so the associations identified should be viewed as preliminary and with caution until more trials are conducted. Second, because dose-response associations are often explored *post hoc* and are not typically the primary focus of trials, there is still a possibility that some failed to report or mention non-significant results. In theory, all studies testing apps for depression and anxiety that gathered objective usage data could have reported the association between engagement and symptoms. Consequently, relevant studies may have been inadvertently excluded from this review, introducing a potential source of reporting bias. Due to the relatively modest number of available studies, we were also not able to apply newer publication bias assessment methods (e.g., PET-PEESE). A final limitation is that we only examined associations between engagement and symptoms. Thus, our results are ultimately only correlational in nature. It is possible that there is at least some degree of reverse causation where individuals who are benefitting more from a mental health app use the app more frequently, rather than the reverse. Future trials that randomly assign participants to varying engagement dosages will be essential to determine whether greater engagement causes improvement.

In conclusion, the current review is the first to synthesize empirical research investigating associations between engagement and clinical outcomes in randomized trials of mental health apps targeting depression and anxiety. We identified preliminary evidence of a small but significant dose-response relationship, which remained stable in various sensitivity analyses that modelled different engagement metrics, symptom outcomes, and app features. However, reporting was highly variable, highlighting the urgent need for the field to establish consensus on which theory-driven engagement metrics should be prioritised in future research. Standardising these metrics will not only enable more rigorous meta-analytic synthesis but also inform the design of mental health apps

that optimise meaningful user engagement and therapeutic outcomes.

Funding

JL supported by a NHMRC Investigator Grant (APP1196948); SBG was supported by the NCCIH (Award No K23AT010879 and R24AT012845).

CRediT authorship contribution statement

Jake Linardon: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **John Torous:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Mariel Messer:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Claudia Liu:** Writing – review & editing, Methodology, Conceptualization. **Imogen H. Bell:** Writing – review & editing, Conceptualization. **Jennifer Nicholas:** Writing – review & editing, Conceptualization. **Simon B. Goldberg:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization.

Declaration of competing interest

None to report

References

- Andersson, G., 2016. Internet-delivered psychological treatments. *Annu. Rev. Clin. Psychol.* 12, 157–179. <https://doi.org/10.1146/annurev-clinpsy-021815-093006>.
- Andersson, G., 2024. Internet-Delivered CBT: Distinctive Features. Taylor & Francis.
- Araya, R., Menezes, P.R., Claro, H.G., Brandt, L.R., Daley, K.L., Quayle, J., Diez-Canseco, F., Peters, T.J., Cruz, D.V., Toyama, M., 2021. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. *JAMA* 325 (18), 1852–1862.
- Bastiaansen, J.A., Ornée, D.A., Meurs, M., Oldehinkel, A.J., 2022. An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). *Psychol. Med.* 52 (13), 2731–2740. <https://doi.org/10.1017/s0033291720004845>. Article Pii s0033291720004845.
- Baumel, A., Muench, F., Edan, S., Kane, J.M., 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J. Med. Internet Res.* 21 (9), e14567.
- Bell, I., Arnold, C., Gilbertson, T., D'Alfonso, S., Castagnini, E., Chen, N., Nicholas, J., O'Sullivan, S., Valentine, L., Alvarez-Jimenez, M., 2023. A personalized, transdiagnostic smartphone intervention (Mello) targeting repetitive negative thinking in young people with depression and Anxiety: pilot randomized controlled trial. *J. Med. Internet Res.* 25, e47860. <https://doi.org/10.2196/47860>.
- Bidargaddi, N., Almirall, D., Murphy, S., Nahum-Shani, I., Kovalcik, M., Pituch, T., Maaieh, H., Strecher, V., 2018. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. *JMIR Mhealth Uhealth* 6 (11), e10123.
- Bradway, M., Gabarron, E., Johansen, M., Zanaboni, P., Jardim, P., Joakimsen, R., Pape-Haugaard, L., Årsand, E., 2020. Methods and measures used to evaluate patient-operated mobile health interventions: scoping literature review. *JMIR Mhealth Uhealth* 8 (4), e16814.
- Catuara-Solarz, S., Skorulski, B., Estella-Aguerrí, I., Avella-García, C.B., Shepherd, S., Stott, E., Hemmings, N.R., de Villa, A.R., Schulze, L., Dix, S., 2022. The efficacy of 'Foundations,' a digital mental health app to improve mental well-being during COVID-19: proof-of-principle randomized controlled trial. *JMIR Mhealth Uhealth* 10 (7), 1–17. <https://doi.org/10.2196/30976>.
- Cohen, J., 1992. A power primer. *Psychol. Bull.* 112, 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Curran, P.J., West, S.G., Finch, J.F., 1996. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychol. Methods* 1 (1), 16.
- Domhardt, M., Mennel, V., Angerer, F., Grund, S., Mayer, A., Büscher, R., Sander, L.B., Cuijpers, P., Terhorst, Y., Baumeister, H., 2025. Processes of change in digital interventions for depression: a meta-analytic review of cognitive and behavioral mediators. *Behav. Res. Ther.*, 104735.
- Donker, T., Cornelisz, I., van Klaveren, C., van Straten, A., Carlbring, P., Cuijpers, P., van Gelder, J.L., 2019. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. *JAMA Psychiatry* 76 (7), 682–690. <https://doi.org/10.1001/jamapsychiatry.2019.0219>.
- Donker, T., Van Klaveren, C., Cornelisz, I., Kok, R.N., Van Gelder, J.L., 2020. Analysis of usage data from a self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized controlled trial. *J. Clin. Med.* 9 (6), 1614.

- Donkin, L., Christensen, H., Naismith, S.L., Neal, B., Hickie, I.B., Glozier, N., 2011. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J. Med. Internet Res.* 13, e52. <https://doi.org/10.2196/jmir.1772>.
- Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- Dwyer, B., Flathers, M., Burns, J., Mikkelsen, J., Perlmutter, E., Chen, K., Ram, N., Torous, J., 2024. Assessing digital phenotyping for app recommendations and sustained engagement: cohort study. *JMIR Format. Res.* 8 (1), e62725.
- Elkes, J., Cro, S., Batchelor, R., O'Connor, S., Yu, L.M., Bell, L., Harris, V., Sin, J., Cornelius, V., 2024. User engagement in clinical trials of digital mental health interventions: a systematic review. *BMC Med. Res. Methodol.* 24 (1), 184.
- Flett, J.A., Fletcher, B.D., Riordan, B.C., Patterson, T., Hayne, H., Conner, T.S., 2019. The peril of self-reported adherence in digital interventions: a brief example. *Intern. Interv.* 18, 100267.
- Forbes, A., Keleher, M.R., Venditto, M., DiBiasi, F., 2023. Assessing patient adherence to and engagement with digital interventions for depression in clinical trials: systematic literature review. *J. Med. Internet Res.* 25, e43727.
- Fortuna, K.L., Brooks, J.M., Umucu, E., Walker, R., Chow, P.I., 2019. Peer support: a human factor to enhance engagement in digital health behavior change interventions. *J. Tech. Behav. Sci.* 4, 152–161.
- Gan, D.Z., McGillivray, L., Han, J., Christensen, H., Torok, M., 2021. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. *Front. Digit. Health* 3, 764079.
- Goldberg, S.B., Baldwin, S.A., Riordan, K.M., Torous, J., Dahl, C.J., Davidson, R.J., Hirshberg, M.J., 2022. Alliance with an unguided smartphone app: validation of the digital working alliance inventory. *Assessment* 29 (6), 1331–1345.
- Goldberg, S.B., Kendall, A.D., Hirshberg, M.J., Dahl, C.J., Nahum-Shani, I., Davidson, R. J., Bray, B.C., 2025. Is dosage of a meditation app associated with changes in psychological distress? It depends on how you ask. *Clin. Psychol. Sci.* 13 (2), 332–349.
- Graham, A.K., Greene, C.J., Kwasny, M.J., Kaiser, S.M., Lieponis, P., Powell, T., Mohr, D. C., 2020. Coached mobile app platform for the treatment of depression and Anxiety among primary care patients A randomized clinical trial. *JAMA Psychiatry* 77 (9), 906–914. <https://doi.org/10.1001/jamapsychiatry.2020.1011>.
- Greer, J.A., Jacobs, J., Pensak, N., MacDonald, J.J., Fuh, C.X., Perez, G.K., Ward, A., Tallen, C., Muzikansky, A., Traeger, L., Penedo, F.J., El-Jawahri, A., Safren, S.A., Pirl, W.F., Temel, J.S., 2019. Randomized trial of a tailored cognitive-behavioral therapy mobile application for anxiety in patients with incurable cancer. *Oncologist* 24 (8), 1111–1120. <https://doi.org/10.1634/theoncologist.2018-0536>.
- Guo, Y., Hong, Y.A., Cai, W., Li, L., Hao, Y., Qiao, J., Xu, Z., Zhang, H., Zeng, C., Liu, C., 2020. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: randomized controlled trial. *J. Med. Internet Res.* 22 (2), e16715.
- Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., Torous, J., 2019. Digital mental health apps and the therapeutic alliance: initial review. *BJPsych Open* 5 (1), e15.
- Higgins, J., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558. <https://doi.org/10.1002/sim.1186>.
- Higgins, J.P., Savovic, J., Page, M.J., Elbers, R.G., Sterne, J.A., 2019. Assessing risk of bias in a randomized trial. *Cochrane Handbook for Systematic Reviews of Interventions*, pp. 205–228.
- Kazdin, A.E., 2017. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. *Behav. Res. Ther.* 88, 7–18. <https://doi.org/10.1016/j.brat.2016.06.004>.
- Kulikov, V.N., Crosthwaite, P.C., Hall, S.A., Flannery, J.E., Strauss, G.S., Vierra, E.M., Koepsell, X.L., Lake, J.I., Padmanabhan, A., 2023. A CBT-based mobile intervention as an adjunct treatment for adolescents with symptoms of depression: a virtual randomized controlled feasibility trial. *Front. Digit. Health* 5, 1062471. <https://doi.org/10.3389/fdgh.2023.1062471>.
- Lacey, C., Frampton, C., Beaglehole, B., 2023. vRcome - self-guided virtual reality for specific phobias: a randomised controlled trial. *Aust. N.Z. J. Psych.* 57 (5), 736–744. <https://doi.org/10.1177/00048674221110779>.
- Li, Y., Guo, Y., Hong, Y.A., Zeng, Y., Monroe-Wise, A., Zeng, C., Zhu, M., Zhang, H., Qiao, J., Xu, Z., 2022. Dose-response effects of patient engagement on health outcomes in an mHealth intervention: secondary analysis of a randomized controlled trial. *JMIR Mhealth Uhealth* 10 (1), e25586.
- Linardon, J., Fuller-Tyszkiewicz, M., 2020. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review [Article]. *J. Consult. Clin. Psychol.* 88 (1), 1–13. <https://doi.org/10.1037/ccp000459>.
- Linardon, J., Fuller-Tyszkiewicz, M., Firth, J., Goldberg, S.B., Anderson, C., McClure, Z., Torous, J., 2024a. Systematic review and meta-analysis of adverse events in clinical trials of mental health apps. *NPJ Digit. Med.* 7 (1), 363. <https://doi.org/10.1038/s41746-024-01388-y>.
- Linardon, J., Torous, J., 2025. Integrating artificial intelligence and smartphone technology to enhance personalized assessment and treatment for eating disorders. *Int. J. Eat. Disorder.* 58 (8), 1415–1424.
- Lipsey, M.W., Wilson, D., 2001. *Practical Meta-Analysis*. Sage Publications.
- Liu, C., Torous, J., Fuller-Tyszkiewicz, M., Messer, M., Anderson, C., Soliman, O.M., Linardon, J., 2025. Uptake, Adherence, and Attrition in Clinical Trials of Depression and Anxiety Apps: A Systematic Review and Meta-Analysis. *JAMA Psychiatry*. <https://doi.org/10.1001/jamapsychiatry.2025.3439>.
- Looyestyn, J., Kermot, J., Boshoff, K., Ryan, J., Edney, S., Maher, C., 2017. Does gamification increase engagement with online programs? A systematic review. *PLoS One* 12 (3), e0173403.
- Mantani, A., Kato, T., Furukawa, T.A., Horikoshi, M., Imai, H., Hiroe, T., Chino, B., Funayama, T., Yonemoto, N., Zhou, Q., et al., 2017. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial [Journal Article; Multicenter Study; Randomized Controlled Trial]. *J. Med. Internet Res.* 19 (11), e373. <https://doi.org/10.2196/jmir.8602>.
- Miklowitz, D.J., Weintraub, M.J., Ichinose, M.C., Denenny, D.M., Walshaw, P.D., Wilkerson, C.A., Frey, S.J., Morgan-Fleming, G.M., Brown, R.D., Merranko, J.A., Arevian, A.C., 2023. A randomized clinical trial of technology-enhanced Family-focused therapy for youth in the early stages of mood disorders. *JAACAP Open* 1 (2), 93–104. <https://doi.org/10.1016/j.jaacop.2023.04.002>.
- Moberg, C., Niles, A., Beermann, D., 2019. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. *J. Med. Internet Res.* 21 (6), e12556.
- Mohr, D.C., Schueller, S.M., Tomasino, K.N., Kaiser, S.M., Alam, N., Karr, C., Vergara, J. L., Gray, E.L., Kwasny, M.J., Lattie, E.G., 2019. Comparison of the effects of coaching and receipt of app recommendations on depression, Anxiety, and engagement in the IntelliCare platform: factorial, randomized controlled trial. *J. Med. Internet Res.* 21 (8), e13609. <https://doi.org/10.2196/13609>.
- Nahum-Shani, I., Shaw, S.D., Carpenter, S.M., Murphy, S.A., Yoon, C., 2022. Engagement in digital interventions. *Am. Psychol.* 77 (7), 836.
- Newman, M.G., Jacobson, N.C., Rackoff, G.N., Bell, M.J., Taylor, C.B., 2021. A randomized controlled trial of a smartphone-based application for the treatment of anxiety. *Psychother. Res.* 31 (4), 443–454. <https://doi.org/10.1080/10503307.2020.1790688>.
- Ng, M.M., Firth, J., Minen, M., Torous, J., 2019. User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatric Services* 70 (7), 538–544.
- Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., 2015. Promoting an open research culture. *Science* 348 (6242), 1422–1425.
- Nwosu, A., Boardman, S., Husain, M.M., Doraiswamy, P.M., 2022. Digital therapeutics for mental health: is attrition the Achilles heel? *Front Psychiatry* 13, 900615.
- Oh, J., Jang, S., Kim, H., Kim, J.J., 2020. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *Int. J. Med. Inform.* 140, 104171. <https://doi.org/10.1016/j.ijmedinf.2020.104171>.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372.
- Peake, E., Miller, I., Flannery, J., Chen, L., Lake, J., Padmanabhan, A., 2024. Preliminary efficacy of a digital intervention for adolescent depression: randomized controlled trial. *J. Med. Internet Res.* 26, e48467.
- Perret, S., Alon, N., Carpenter-Song, E., Myrick, K., Thompson, K., Li, S., Sharma, K., Torous, J., 2023. Standardising the role of a digital navigator in behavioural health: a systematic review. *The Lancet Digital Health* 5 (12), e925–e932.
- Peterson, R.A., Brown, S.P., 2005. On the use of beta coefficients in meta-analysis. *J. Appl. Psychol.* 90 (1), 175–181.
- Raevuori, A., Vahlberg, T., Korhonen, T., Hilgert, O., Aittakumpu-Hyden, R., Forman-Hoffman, V., 2021. A therapist-guided smartphone app for major depression in young adults: a randomized clinical trial. *J. Affect. Disord.* 286, 228–238. <https://doi.org/10.1016/j.jad.2021.02.007>.
- Roy, A., Hoge, E.A., Abrante, P., Druker, S., Liu, T., Brewer, J.A., 2021. Clinical efficacy and psychological mechanisms of an app-based digital therapeutic for generalized anxiety disorder: randomized controlled trial. *J. Med. Internet Res.* 23 (12), e26987.
- Santomauro, D.F., Herrera, A.M.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., Abbafati, C., Adolph, C., Amlag, J.O., Aravkin, A.Y., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet* 398 (10312), 1700–1712.
- Saulnier, K.G., Koscinski, B., Flynt, S., Accorso, C., Allan, N.P., 2023. Brief observable anxiety sensitivity treatment: intervention development and a pilot randomized-controlled acceptability and feasibility trial to evaluate a brief intervention for anxiety sensitivity social concerns. *Cogn. Behav. Ther.* <https://doi.org/10.1080/16506073.2023.2288551>.
- Schwob, J.T., Newman, M.G., 2023. Brief imaginal exposure exercises for social anxiety disorder: a randomized controlled trial of a self-help momentary intervention app. *J. Anxiety Disord.* 98, 102749. <https://doi.org/10.1016/j.janxdis.2023.102749>.
- Sirbu, V., David, O.A., Sanchez-Lopez, A., Blanco, I., 2025. Comparative efficacy of PsyPills and OCAT mobile psychological interventions in reducing depressive, anxiety and stress symptoms: a blinded randomized clinical trial. *J. Affect. Disord.* 369, 945–953.
- Six, S.G., Byrne, K.A., Aly, H., Harris, M.W., 2022. The effect of mental health app customization on depressive symptoms in college students: randomized controlled trial. *JMIR Ment. Health* 9 (8), e39516. <https://doi.org/10.2196/39516>.
- Stanley, T.D., 2017. Limitations of PET-PEESE and other meta-analysis methods. *Soc. Psychol. Personal Sci.* 8 (5), 581–591.
- Stiles-Shields, C., Montague, E., Kwasny, M.J., Mohr, D.C., 2018. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. *Psychol. Serv.* <https://doi.org/10.1037/ser0000261>.
- Stiles-Shields, C., Reyes, K.M., Lakhtakia, T., Smith, S.R., Barnas, O.E., Gray, E.L., Krause, C.J., Kruzan, K.P., Kwasny, M.J., Mir, Z., Panjwani, S., Rothschild, S.K., Sánchez-Johnsen, L., Winquist, N.W., Lattie, E.G., Allen, N.B., Reddy, M., Mohr, D. C., 2024a. A personal sensing technology enabled service versus a digital psychoeducation control for primary care patients with depression and anxiety: a pilot randomized controlled trial. *BMC Psychiatry* 24 (1), 1–15. <https://doi.org/10.1186/s12888-024-06284-z>.

- Stolz, T., Schulz, A., Krieger, T., Vincent, A., Urech, A., Moser, C., Westermann, S., Berger, T., 2018. A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. *J. Consult. Clin. Psychol.* 86 (6), 493–504. <https://doi.org/10.1037/ccp0000301>.
- Tighe, J., Shand, F., Ridani, R., Mackinnon, A., De La Mata, N., Christensen, H., 2017. Ibbobly mobile health intervention for suicide prevention in Australian indigenous youth: a pilot randomised controlled trial. *BMJ Open* 7, e013518. <https://doi.org/10.1136/bmjopen-2016013518>.
- Torous, J., Bucci, S., Bell, I.H., Kessing, L.V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A.F., Keshavan, M., Linardon, J., Firth, J., 2021. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality [Article]. *World Psychiatry* 20 (3), 318–335. <https://doi.org/10.1002/wps.20883>.
- Torous, J., Linardon, J., Goldberg, S.B., Sun, S., Bell, I., Nicholas, J., Hassan, L., Hua, N., Milton, A., Firth, J., 2025. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry* 24 (2), 1–19. <https://doi.org/10.1002/wps.21299>.
- Torous, J., Lipschitz, J., Ng, M., Firth, J., 2020. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J. Affect. Disord.* 263, 413–419.
- Torous, J., Nicholas, J., Larsen, M.E., Firth, J., Christensen, H., 2018. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid. Based Ment. Health* 21, 116–119. <https://doi.org/10.1136/eb-2018-102891>.
- Torous, J., Wisniewski, H., Bird, B., Carpenter, E., David, G., Elejalde, E., Fulford, D., Guimond, S., Hays, R., Henson, P., 2019. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *J. Tech. Behav. Sci.* 4, 73–85.
- World Health Organization, 2022. *World Mental Health Report: Transforming Mental Health for All*. WHO.
- Yatziv, S.L., Pedrelli, P., Baror, S., DeCaro, S.A., Shachar, N., Sofer, B., Hull, S., Curtiss, J., Bar, M., 2024. Facilitating thought progression to reduce depressive symptoms: randomized controlled trial. *J. Med. Internet Res.* 26, e56201. <https://doi.org/10.2196/56201>.
- Zuccolo, P.F., Brunoni, A.R., Borja, T., Matijasevich, A., Polanczyk, G.V., Fatori, D., 2024. Efficacy of a standalone smartphone application to treat postnatal depression: a randomized controlled trial. *Psychother. Psychosom.* 93 (6), 412–424. <https://doi.org/10.1159/000541311>.