The efficacy of mindfulness apps on symptoms of depression and anxiety: An updated meta-analysis of randomized controlled trials

Jake Linardon [1,2]

Mariel Messer [1]

Simon B. Goldberg [3,4]

Matthew Fuller-Tyszkiewicz [1,2]

[1] School of Psychology, Deakin University, Geelong, Victoria, Australia

[2] Center for Social and Early Emotional Development, Deakin University, Burwood, Victoria, Australia

[3] Department of Counselling Psychology, University of Wisconsin – Madison, Madison, WI, USA

[4] Centre for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

Correspondence: Jake.Linardon@deakin.edu.au, 1 Gheringhap St, Geelong VIC 3220

Abstract

Mindfulness apps have become popular tools for addressing symptoms of depression and anxiety. Since the publication of earlier meta-analyses evaluating the efficacy of mindfulness apps for depression and anxiety symptoms, over 20 randomized controlled trials (RCTs) have been conducted. There is a need for an updated meta-analysis that quantifies the effects of mindfulness apps on these symptoms and tests for potential moderators.. Random effects meta-analyses were conducted on 43 RCTs. Small, significant effect sizes were found for symptoms of depression (N_{comp}=46, N=5852, g=0.24, 95% CI=0.17, 0.31, NNT = 13.57) and anxiety (N_{comp}=48, N=6082, g=0.28, 95% CI=0.21, 0.35, NNT = 11.47) in favour of mindfulness apps over control groups. This effect was not explained by symptom deterioration in participants allocated to control groups. Effects remained stable when restricting analyses to lower risk of bias and larger sample trials. No significant moderators were observed, except trials that offered monetary compensation produced larger effects on depression. Non-significant effects were observed when comparing mindfulness apps to active therapeutic comparisons (g=-0.15 depression, g=0.10 anxiety), though the number of studies was low. Growing evidence indicates that mindfulness apps can acutely reduce symptoms of depression and anxiety, although higher quality studies with longer follow-ups are needed.

Keywords: mindfulness; smartphone apps; depression; anxiety; meta-analysis; randomized controlled trial

Introduction

Depressive and anxiety disorders are serious mental health disorders affecting a significant proportion of the global population (Vos et al., 2016). The two disorders overlap substantially, with more than 8 in 10 people suffering from an anxiety disorder also experiencing a depressive disorder in their lifetime (Lamers et al., 2011). Both depression and anxiety are associated with chronic physical comorbidities, poor social functioning, substance use and abuse, and increased rates of suicidality (Miret et al., 2013; Wittchen, 2002). They also place an enormous strain on the healthcare system and contribute to losses in worldwide work participation and productivity (Chisholm et al., 2016; Greenberg et al., 2015; Santomauro et al., 2021).

Mindfulness-based interventions (MBIs; Kabat-Zinn, 2003) have become a popular approach for addressing depressive and anxiety symptoms. MBIs teach the person how to orient attention towards present moment experiences in a non-judgemental, curious, and accepting manner via formal meditation (e.g., breath awareness, body scans) and other informal activities (e.g., bringing non-judgmental attention to daily activities). This is thought to disrupt many of the processes that contribute to the onset and maintenance of emotional disorders, such as rumination, experiential avoidance, cognitive interpretation bias, and emotion dysregulation (Beck & Bredemeier, 2016; Craske, 1999). In the context of depressive and anxiety symptoms, hundreds of randomized controlled trials (RCTs) have evaluated MBIs, mostly delivered in group-based facilitator-led settings, but an increasing number of trials have tested MBIs delivered through web programs and mobile apps (Galante et al., 2023). A recent umbrella review of meta-analyses confirmed the efficacy of

traditional, group-based MBIs relative to control conditions (Goldberg, Riordan, et al., 2022). [1]

Most MBIs evaluated in controlled research settings have been administered in a face-to-face format. This is because standardized MBIs like Mindfulness-Based Stress Reduction (MBSR; Kabat-Zinn, 2007) and Mindfulness-Based Cognitive Therapy (MBCT; Segal et al., 2018) are traditionally delivered following a written curriculum in facilitator-led, group-based courses over eight weekly sessions. However, facilitator-led, face-to-face interventions like these have limited accessibility due to finite instructor availability, program costs, and logistical challenges (Torous et al., 2021). One proposed solution to addressing these barriers and increasing the availability of MBIs is to translate content for delivery via smartphone applications ("apps").

App-based interventions have obvious advantages over in-person delivery in terms of cost, scalability, and anonymity (Goldberg, Lam, et al., 2022). Unlike traditional in-person delivery, apps also enable access to therapeutic strategies anytime and anywhere, making it possible to tailor the type, intensity, and format of content based on passive (e.g., GPS coordinates) and active (e.g., symptom level) data (Linardon et al., 2019; Torous et al., 2021). Since most people own a smartphone and take it with them wherever they go, users are able to practice key skills repeatedly throughout the day to prevent symptom onset or exacerbation (Bakker et al., 2016).

The potential advantages of app-based interventions have resulted in rapid development of a large number of mindfulness apps. To date, there are nearly 300 mindfulness apps available for download, making them one of the most common types

[1] We define control conditions as either passive (waitlist, assessment only) or an active (resources that control for time and expectations) and differentiate them from active comparisons (interventions intended to have therapeutic value, such as a different treatment approach [CBT]).

of apps for enhancing mental health and wellbeing (Camacho et al., 2022). For example, the Calm app reports over 40 million downloads and over 1 million paid subscribers (Gebel, 2019), while the Headspace app reports 70 million downloads and over 2 million paid subscribers (Headspace, 2021). Mindfulness apps attempt to include the core content delivered in traditional in-person MBIs, namely instruction around effective practice of mindfulness meditation (Segal et al., 2018). However, unlike traditional MBIs, mindfulness apps are mostly delivered in a self-guided format, meaning that users do not have the opportunity to interact with fellow group members or a mindfulness instructor (Goldberg et al., 2020). Furthermore, in-person MBIs typically cover basic elements of cognitive therapy (Segal et al., 2018), which are not features of existing mindfulness apps tested in research settings. Despite widespread accessibility of apps of this kind, most commercially available mindfulness apps lack evidence from controlled clinical research to substantiate claims made about their benefits (e.g., in advertising). This combination of widespread use in the absence of definitive evidence has prompted a call for enhancing the evidence-base for mindfulness apps, leading to the proliferation of RCTs in recent years evaluating the viability and clinical utility of mindfulness apps for depression and anxiety (Gál et al., 2021).

Two prior meta-analyses have synthesized results from RCTs of mindfulness meditation apps on symptoms of depression and anxiety (Gál et al., 2021; Tan et al., 2022). Both found mindfulness apps to outperform control conditions on symptoms of depression and anxiety, with Gál et al. (2021) reporting a pooled effect size of $g = 0.33$ for depression and $g = 0.28$ for anxiety based on 15 trials each, and Tan et al. (2022) reporting a pooled effect size of $g = 0.21$ for depression and $g = 0.08$ for anxiety based on eight and seven trials, respectively.

Since the publication of these meta-analyses, which included data from RCTs conducted prior to December 2020, more than 20 trials of mindfulness apps have been published, highlighting the need to provide a more up-to-date synthesis of this field. Additionally, the large number of trials now enables examination of potential moderating variables, which is important for understanding the circumstances under which mindfulness apps are most or least effective. Prior meta-analyses on digitally-delivered interventions may shed light on potentially important moderators in this context (Firth et al., 2017; Heber et al., 2017; Linardon et al., 2019). The effects of mindfulness apps are likely larger in trials that deliver a passive versus placebo control, in light of prior assertions that some of the therapeutic effects of mental health apps are explained by use of the digital device itself rather than core elements of the intervention (i.e., “digital placebo”; Torous & Firth, 2016). Similarly, effects may be larger in samples with versus without pre-existing mental health problems, as the former subgroup (Linardon et al., 2020). There is also evidence that shorter follow-up durations produce larger effect size estimates than longer follow-ups in web-based interventions (Heber et al., 2017), which may be explained by the fact that it is more difficult to keep participants engaged in interventions over longer periods. Finally, trials that require contact with the researcher (vs. fully remote trials with no researcher contact) and provide participants monetary compensation may produce larger effects, as both factors have been reliably associated with sustained engagement (Linardon, 2023; Linardon & Fuller-Tyszkiewicz, 2020), which might in turn produce greater rate of symptom change.

Another reason for the need to provide an updated synthesis is that the type, nature, and functionality of mindfulness apps have likely evolved at a rapid pace in recent years, potentially affecting effect size estimates. Recent work shows that 1 in 5

commercially available mental health apps receive updates every three months, with changes to privacy policy, costs, functionality, and delivery of engagement features cited as the most frequent updates (Stoeckl et al., 2023). However, frequent updates of this nature are unlikely to represent those mindfulness apps that are developed specifically for research purposes; commercially available apps typically have substantially larger budgets, and are consequently more visually aesthetic and offer additional innovative features (e.g., paid subscription options), despite both displaying similar content and principles related to mindfulness practice (Lagan et al., 2021). For these reasons, it is timely to conduct an updated meta-analysis evaluating the efficacy of mindfulness apps on symptoms of depression and anxiety so that practitioners, policy makers, and the public have up-to-date knowledge on the evidence base and clinical utility of apps of this kind.

The aims of this updated meta-analysis are twofold: first, to evaluate the effects of mindfulness apps on symptoms of depression and anxiety reported in RCTs. Second, to examine whether various study-level factors are associated with effect sizes.

Method

Search Strategy and Study Selection

We searched four online databases (PsycInfo, Medline, Web of Science, and ProQuest database for dissertations) on the 24th January 2023 (updated on the 18th October 2023) using the terms (mindful* OR meditate*) AND (app-based OR application OR mobile OR phone OR smartphone OR app-delivered OR mhealth OR m-health OR mobile-based OR mobile-health OR cellphone* OR iphone OR android) AND (Random* OR trial OR RCT). The secondary search strategy involved searching through the reference lists of included studies, relevant reviews on mental health

apps, and clinical trials registries in case of any studies (published or unpublished) not captured by the primary search strategy. This review was pre-registered (CRD42023414171) and adhered to the PRIMSA guidelines (Moher et al., 2009).

We included RCTs of MBIs delivered via a smartphone app that were evaluated against either a control condition or an active comparison and assessed symptoms of depression or anxiety as an outcome. No sample restrictions were applied. Published and unpublished studies were eligible for inclusion. Unpublished studies were searched in three ways: (1) using the search terms in the ProQuest Database for Dissertations; (2) searching through reference lists of prior reviews on mental health apps that included unpublished trials; and (3) searching through clinical trials registries (Australian New Zealand Clinical Trials Registry, ClinicalTrials.gov) for studies near completed but not published. Mindfulness had to be the central component of the app, which we defined on the basis of criteria put forth by Crane et al. (2017). Specifically, the mindfulness app needed to emphasize formal meditation practice; interventions that merely encourage an attitude of mindfulness ("mindfulness-informed" approaches) without emphasis on meditation practice ("mindfulness-based" approaches), such as acceptance and commitment therapy and dialectical behaviour therapy approaches, were excluded. Adjunctive interventions (e.g., mindfulness app plus face-to-face therapy) were also excluded. No restrictions were placed on the type of comparison condition delivered. Control groups were either categorized as inactive (i.e., waitlist, assessment-only) or active (information resources, non-therapeutic app, music listening, care as usual etc.). Active psychological comparisons were those that were intended to be therapeutic, such as a CBT app or face-to-face counselling sessions. 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.

Data Extraction

A systematic extraction process was applied to retrieve the following information from eligible trials: authors, sample characteristics, mindfulness app, comparison group, sample size, length of assessment, extent of researcher contact, whether monetary compensation was offered, recommended practice of mindfulness skills throughout the study, risk of bias indicators, and outcome measures. Extraction was performed by two authors (JL & MM).

Risk of Bias

The Cochrane Collaboration Risk of Bias (RoB) tool was used to assess for risk of bias (Higgins & Green, 2011). The following five domains were rated for each trial: random sequence generation, allocation concealment, blinding of participants or personnel, blinding of outcome assessment, and incomplete outcome data. Each domain was rated as either 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 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). Attrition bias was rated as low if the trial authors included all randomized participants in their analyses (i.e., they adhered to the intention-to-treat principle).

Meta-Analyses

All analyses were conducted using Comprehensive Meta-Analysis Version 3.0 (Borenstein et al., 2009). For each comparison between the mindfulness 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. The standardized mean difference was then converted to Hedges' *g* to correct for small sample bias, which is relevant given noted low sample size for many studies in the earlier reviews (Gál et al., 2021; Tan et al., 2022). If means and standard deviations were not reported, effect sizes were calculated from other reported statistics (e.g., *t*, or *p*-values for group comparison) using conversion equations (Lipsey & Wilson, 2001). To calculate a pooled effect size, each study's effect size was weighted by its inverse variance. A positive *g* indicates that the mindfulness app had lower symptoms of depression and anxiety 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 small (Cohen, 1992). If data from both intention-to-treat and completer analyses were presented, the former were extracted and analyzed.

While Hedges' *g* attempts to standardize post-treatment intervention effects across studies, it can provide an incomplete picture of efficacy since it may reflect a combination of improvements in the intervention group and/or deterioration in the control condition. Consequently, we supplemented these between-group Hedges' *g* estimates with estimates of within-group effects for control and intervention groups. These estimates were calculated separately for control and intervention groups using the method proposed by Abrams et al. (2005) to calculate a standard deviation (*SD*) value for change to divide change in means by:

$$\sigma_{change} = sqrt[\sigma_{pre}^2 + \sigma_{post}^2 - (2\rho \times \sigma_{pre} \times \sigma_{post})]$$

where σ = standard deviation, sqrt = square root, and *p* = estimate of correlation between baseline and post-intervention scores within group. Given the correlation value is not typically reported in RCTs, we estimated across a range of values (*r* = .1., .2, … .9). This full set of results is provided in a supplementary file (Tables S1-S8), but are summarized in the Results section.

We also supplemented between-group effect size estimates with estimates of NNT to convey the practical impact of the weighted-mean for intervention effects, using an online calculator (Magnusson, 2022). 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 group.

Since we expected considerable heterogeneity among the studies, random effects models were employed (Borenstein et al., 2009). 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 variance (Higgins & Thompson, 2002). Subgroup analyses were conducted to explore sources of heterogeneity under a mixed effects model, which pools studies within a subgroup using a random effects model, but tests for significant differences between subgroups using fixed effects models. Small study bias was also examined through the trim-and-fill procedure (Duval & Tweedie, 2000).

Results

Study Characteristics

Figure 1 presents the PRISMA flowchart of the literature search. Forty-three papers (45 studies) met full inclusion criteria; only 3 studies were unpublished (all dissertations). Eligible studies mostly used an unselected sample, defined as a sample of individuals who were not screened for – or required to exhibit – the presence or

absence of mental health problems, such as students, general population of adults, and employees. Other samples studied were cancer, obstetric, myeloproliferative neoplasm, and intensive care unit patients. Fourteen studies sampled individuals with pre-existing mental health problems, including elevated depression, anxiety, and stress. The most common mindfulness app delivered was Headspace (k = 15) followed by Calm (k = 5). The type of comparison condition varied; 22 studies used an inactive control condition (waitlist), and 22 used an active control (e.g., information resources, a non-therapeutic app, music listening activities, online math training, care as usual) that controlled for time, attention, and participant expectations. Only four trials employed an active therapeutic comparison, which included a face-to-face intervention (k = 2), telephone-delivered counselling (k = 1), and a behavioral activation app (k = 1). Twenty-one trials were fully remote and did not require any contact with the researchers (those that did mostly involved the researcher meeting with the participant either in-person, or via zoom or telephone to determine their eligibility or provide instructions for app use), and 14 offered participants monetary compensation for completing study assessments. Length of post-test assessment ranged from 10 days to 8 weeks. See Table 1 for further details about characteristics of included studies.

Risk of bias domain ratings for each study is presented in Supplementary Materials Table 1. All studies used self-report scales to assess symptoms of depression and anxiety. A total of 29 studies used an adequate sequence generation for randomization, 10 satisfied criteria for allocation concealment, 17 were rated as low risk for blinding of participants, and 22 conducted analyses based on the intention-to-treat principle. Only four studies (8.8%) satisfied all five risk of bias criteria, six

(13.3%) satisfied four criteria, 15 satisfied three criteria (33.3%), 14 (31.1%) satisfied two criteria, and five (11.1%) satisfied only one criterion.

Of note, one trial each that was excluded in this meta-analysis were included in the earlier meta-analysis by Gál et al. (2021) and Tan et al. (2022). The first (Moberg et al., 2019) was excluded because the app tested (Pacifica) was mostly based on CBT principles, with mindfulness not being the central component. The second (McClain, 2017) was excluded because the mindfulness exercises were delivered via text messages, not a smartphone app.

Mindfulness Apps Versus Control Conditions

Depressive Symptoms

The pooled effect size for the 46 comparisons (N = 5852) between mindfulness apps and control conditions (passive and active) on depressive symptoms was a small but statistically significant g = 0.24 (95% CI = 0.17, 0.31, NNT = 13.57). Statistical heterogeneity was low (I^2 = 32%). The pooled effect size was the same when applying the trim-and-fill procedure, and comparable when restricting the analyses to lower risk of bias (g = 0.29, NNT = 11.04) and larger sample (g = 0.31, NNT = 10.26) trials.

In the previous analyses, we included a few trials in which more than one mindfulness app condition was compared with the same control condition (or vice versa). These comparisons were not independent from each other, which may have artificially reduced the heterogeneity estimate and affected the pooled effect size. To deal with this, we ran sensitivity analyses in which the comparison with the smallest effect size was only included in the analysis, and then repeated this again for the comparison with the largest effect size. These sensitivity analyses ensured that only one comparison per study was included in the meta-analysis. These sensitivity

analyses yielded a pooled effect size very similar to the overall effect, as can be seen in Table 2.

Across studies with available pre- and post-intervention data, the between-group effect estimate was larger than the average within-group effect for the intervention participants in only three studies (9% of available studies; see Supplementary Tables S2, S6-S8 for further details), with differences within 0.19 standard deviations of each other. This suggests that effects are unlikely driven by deterioration in the control group.

Anxiety Symptoms

The pooled effect size for the 48 comparisons (N = 6082) between mindfulness apps and control conditions (passive and active) on anxiety symptoms was a small but statistically significant g = 0.28 (95% CI = 0.21, 0.35, NNT = 11.47), with moderate heterogeneity (I^2 = 44%). Effect sizes remained significant and similar in magnitude when applying the trim-and-fill procedure, and when restricting the analyses to one comparison per study, low risk of bias trials, and larger sample trials (see Table 2). Across studies with available pre- and post-intervention data, the between-group effect estimate was larger than the average within-group effect for the intervention participants in five studies (14% of available studies; see Supplementary Tables S1, S3-S5 for further details), with modest differences within 0.13 standard deviations of each other.

Subgroup Analyses. Results from the subgroup analyses are also presented in Table 2. One significant moderation effect emerged. Trials that offered participants monetary compensation were associated with larger effect size estimates for depressive symptoms compared to trials that did not offer any monetary

compensation. Neither sample type, length of follow-up, mindfulness app, type of control group, nor provision of researcher contact were associated with effect sizes.

Mindfulness Apps Versus Active Therapeutic Comparisons

The pooled effect sizes for the comparison between mindfulness apps and active therapeutic comparisons for depressive (N_{comp} = 3, N = 181, g = -0.15, NNT = 22.43) and anxiety (N_{comp} = 4, N = 235, g = 0.10, NNT = 34.30) was small and non-significant, although the number of studies was low. See Table 2 for results of these analyses.

Discussion

We report on an updated meta-analysis evaluating the efficacy of mindfulness apps for symptoms of depression and anxiety. Not only do we include a much larger number of trials than the two earlier meta-analyses on mindfulness apps (Gál et al., 2021; Tan et al., 2022), but we extend from their syntheses in three important and unique ways. First, by formally testing whether certain study, sample, and intervention characteristics moderate effect size estimates. Second, by reporting other more clinically meaningful metrics (number-needed-to-treat) than the standardized mean difference to aid interpretation of the effects of mindfulness apps. Third, by investigating whether the positive effects found in these trials are explained by either improvements reported by participants allocated to the mindfulness app or deteriorations reported by participants in the control group (or some combination of these).

Overall effect sizes of g = 0.24 and 0.28 were respectively observed for symptoms of depression and anxiety, which are similar in magnitude to what was reported in the meta-analysis by Gál et al. (2021) and were not explained by deteriorations in symptoms reported by participants in the control group. These

estimates are also comparable to the most recent effect sizes reported for CBT apps on symptoms of depression (g = 0.35; 95% CI =0.28, 0.42) and anxiety (g = 0.30; 95% CI =0.24, 0.36), suggesting that all types of mental health apps may produce only small benefits on these problems (Linardon et al., in press). Per NNT estimates, these weighted effect sizes suggest that more than 10 participants would need to be treated for one individual to see symptom improvements relative to control group participants. This pattern of effects remained stable when restricting the analyses to lower risk of bias and larger sample trials. We found little evidence that study characteristics were associated with effect sizes. Findings overall suggest that stand-alone mindfulness apps may have small but positive effect on improving symptoms of depression and anxiety relative to control conditions, but also highlight areas for further research exploration.

One unexpected finding was that no sample, trial, or intervention characteristics were associated with effect sizes, with the exception of offering participant monetary compensation for depressive symptoms (which was likely a spurious finding given the number of tests performed). Previous meta-analyses on mental health apps have found that trials incorporating inactive control conditions produce larger effect sizes than trials using an active or placebo control (Firth et al., 2017; Linardon, 2020). We failed to identify such an association, which may be explained by our method of categorization. In particular, active controls involved a combination of different conditions, including care as usual, information resources, and placebo apps. Combining different conditions like these may have masked any subgroup effects, as certain control conditions may contribute more or less to different placebo effects (i.e., non-therapeutic apps vs. educational resources). Unfortunately we were not able to further categorize different “active control” groups to empirically test this.

We also found no evidence that the presence versus absence of researcher contact was associated with effect sizes. This variable has been shown to be strongly associated with attrition in app trials (Linardon, 2023; Linardon & Fuller-Tyszkiewicz, 2020), and while researcher contact may motivate participants to remain in app-based trials and complete its follow-up assessments, perhaps it does not lead to more sustained app usage and, consequently, greater symptom reduction. It is possible that there are other characteristics associated with effect sizes not tested in this study, highlighting the need for future research to uncover the specific conditions under which mindfulness apps are most (or least) effective. Alternatively, the moderate heterogeneity may suggest that self-guided mindfulness apps are only moderately effective.

Only four studies compared a mindfulness app to an active intervention, none of which were powered to detect small between-group effects (*n*s ranged from 20–51 per condition). Here we found negligible, non-significant effect sizes (N = 181 and 235 for analyses on depression and anxiety, respectively), which aligns with prior meta-analyses on web-based interventions (Carlbring et al., 2018). There is thus a need for further studies to be designed to test for equivalence between mindfulness apps and either face-to-face MBIs or other types of mental health apps. Adequately powered equivalence trials like these may also help to uncover moderators of response, which could provide crucial knowledge about which intervention works best, for whom, and under what conditions, thereby informing more personalized models of mental health care (Kraemer et al., 2002).

There are important limitations to this meta-analysis that must be considered. First, effects on anxiety and depressive symptoms were only calculated at post-test given the dearth of studies conducting longer-term follow-up assessment. Thus,

whether the positive effects of mindfulness apps on symptoms of depression and anxiety remain stable over longer periods remains an open question worthy of future investigation. Second, risk of bias was considerable in many of the included trials. Although effects remained stable when restricting the analyses to low risk of bias trials, the present effect estimates should be considered with a degree of caution. This is particularly true of studies that delivered the Headspace app, as recent concerns have been raised about the potential for conflicts of interest and its impact on efficacy estimates (O'Daffer et al., 2022). Third, most included trials sampled non-clinical populations, such as students or adults from the general population. The generalizability of findings to individuals with either a diagnosed depressive or anxiety disorder or scoring above a clinical cut-off on screening measures remains limited. Fourth, very few trials reported outcomes pertaining to remission, reliable change, or deterioration, meaning that we had to exclusively rely on analysing symptom change. However, clinicians and end-users may want to know how many individuals recover from these symptoms after using a mindfulness app.

Findings highlight possible ways in which these apps may be incorporated within models of mental health care. Perhaps mindfulness apps could be situated within the stepped-care framework, in which scalable, low intensity, and inexpensive self-management tools like these are offered initially, with more intensive resources reserved for those who fail to benefit after a certain time-period (van Straten et al., 2015). This might be important in educational settings, where mental health resources are lacking (Harrer et al., 2019). Alternatively, mindfulness apps could be recommended to individuals placed on a waiting list for counselling services as a way to keep the user engaged, maintain motivation, build confidence in ability to change, and alleviate certain symptoms at least to a modest degree. Another way to embed

mindfulness apps into traditional care could be for therapists to encourage their use between sessions so that clients can more regularly and efficiently practice key skills in moments of need. There is an urgent need for future research to establish practical, feasible and appropriate ways for implementing mindfulness apps within healthcare systems.

In conclusion, the rapidly growing evidence to date indicates that mindfulness apps can reduce symptoms of depression and anxiety in the short-term, but whether effects are sustained over longer periods remains unclear. Growing evidence suggests that mindfulness apps may be particularly useful for asymptomatic or at risk populations. However, it is important to acknowledge that, alone, they are unlikely to adequately address the mental health needs at the population level and other forms of psychological and pharmacological treatment are still required to produce larger effects on mental health symptoms. Nevertheless, mindfulness apps may be useful for producing short-term symptom relief. Developers of future mindfulness apps may benefit from incorporating technological innovations that may bolster their effectiveness, such as using passively collected data (GPS location, physiological changes) to deliver tailored interventions or allowing interactions with digital conversational agents that can provide in-the-moment support.

References

Abbott, D. (2018). *Evaluating a Smartphone Mindfulness Intervention's Effectiveness at Reducing Anxiety and Worry.* (M.A.), University of Central Oklahoma, United States -- Oklahoma. ProQuest Dissertations & Theses Global database. (10813715)

Abrams, K. R., Gillies, C. L., & Lambert, P. C. (2005). Meta-analysis of heterogeneously reported trials assessing change from baseline. *Statistics in Medicine, 24*, 3823-3844.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Mental Health, 3*. doi:10.2196/mental.4984

Bear, K. A., Barber, C. C., & Medvedev, O. N. (2022). The Impact of a Mindfulness App on Postnatal Distress. *Mindfulness, 13*, 2765-2776. doi:10.1007/s12671-022-01992-7

Beck, A. T., & Bredemeier, K. (2016). A unified model of depression: Integrating clinical, cognitive, biological, and evolutionary perspectives. *Clinical Psychological Science, 4*, 596-619.

Bhayee, S., Tomaszewski, P., Lee, D. H., Moffat, G., Pino, L., Moreno, S., & Farb, N. A. (2016). Attentional and affective consequences of technology supported mindfulness training: a randomised, active control, efficacy trial. *BMC Psychology, 4*, 1-14.

Boden, L. M., Rodriguez, C., Kelly IV, J. D., Khalsa, A. S., & Casper, D. S. (2023). Mindfulness Applications: Can They Serve as a Stress, Anxiety, and Burnout Reduction Tool in Orthopaedic Surgery Training? A Randomized Control Trial. *JBJS Open Access, 8*, e22.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*: John Wiley & Sons.

Borjalilu, S., Mazaheri, M. A., & Talebpour, A. (2019). Effectiveness of Mindfulness-Based Stress Management in The Mental Health of Iranian University Students: A Comparison of Blended Therapy, Face-to-Face Sessions, and mHealth App (Aramgar). *Iranian Journal of Psychiatry and Behavioral Sciences, 13*. doi:10.5812/ijpbs.84726

Bosso, K. B. (2020). *The effects of mindfulness training on BDNF levels, depression, anxiety, and stress levels of college students.* (81), ProQuest Information & Learning.

Bostock, S., Crosswell, A. D., Prather, A. A., & Steptoe, A. (2019). Mindfulness On-The-Go: Effects of a Mindfulness Meditation App on Work Stress and Well-Being. *Journal of Occupational Health Psychology, 24*, 127-138. doi:10.1037/ocp0000118

Camacho, E., Cohen, A., & Torous, J. (2022). Assessment of Mental Health Services Available Through Smartphone Apps. *JAMA Network Open, 5*, e2248784-e2248784.

Carissoli, C., Villani, D., & Riva, G. (2015). Does a Meditation Protocol Supported by a Mobile Application Help People Reduce Stress? Suggestions from a Controlled Pragmatic Trial. *Cyberpsychology Behavior and Social Networking, 18*, 46-53. doi:10.1089/cyber.2014.0062

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cognitive Behaviour Therapy, 47*, 1-18.

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: a global return on investment analysis. *The Lancet Psychiatry, 3*, 415-424.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi:10.1037/0033-2909.112.1.155

Cox, C. E., Hough, C. L., Jones, D. M., Ungar, A., Reagan, W., Key, M. D., . . . Porter, L. S. (2019). Effects of mindfulness training programmes delivered by a self-directed mobile app and by telephone compared with an education programme for survivors of critical illness: a pilot randomised clinical trial. *Thorax, 74*, 33-42. doi:10.1136/thoraxjnl-2017-211264

Crane, R. S., Brewer, J., Feldman, C., Kabat-Zinn, J., Santorelli, S., Williams, J. M. G., & Kuyken, W. (2017). What defines mindfulness-based programs? The warp and the weft. *Psychological Medicine, 47*, 990-999.

Craske, M. G. (1999). *Anxiety disorders: psychological approaches to theory and treatment*: Westview Press.

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. doi:10.1111/j.0006-341X.2000.00455.x

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry, 16*, 287-298. doi:10.1002/wps.20472

Fish, M. T., & Saul, A. D. (2019). The Gamification of Meditation: A Randomized-Controlled Study of a Prescribed Mobile Mindfulness Meditation Application in

Reducing College Students' Depression. *Simulation & Gaming, 50*, 419-435. doi:10.1177/1046878119851821

Flett, J. A. M., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile Mindfulness Meditation: a Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health. *Mindfulness, 10*, 863-876. doi:10.1007/s12671-018-1050-9

Forbes, G., Newton, S., Calvete, C. C., Birch, J., Dodds, J., Steed, L., . . . Taylor, S. (2020). MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. *BMJ Open, 10*, e030164.

Gál, É., Ștefan, S., & Cristea, I. A. (2021). The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. *Journal Of Affective Disorders, 279*, 131-142. doi:https://doi.org/10.1016/j.jad.2020.09.134

Galante, J., Friedrich, C., Dalgleish, T., Jones, P. B., & White, I. R. (2023). Systematic review and individual participant data meta-analysis of randomized controlled trials assessing mindfulness-based programs for mental health promotion. *Nature Mental Health, 1*, 462-476.

Gao, M., Roy, A., Deluty, A., Sharkey, K. M., Hoge, E. A., Liu, T., & Brewer, J. A. (2022). Targeting anxiety to improve sleep disturbance: A randomized clinical trial of app-based mindfulness training. *Psychosomatic Medicine, 84*, 632-642.

Goldberg, S. B., Imhoff-Smith, T., Bolt, D. M., Wilson-Mendenhall, C. D., Dahl, C. J., Davidson, R. J., & Rosenkranz, M. A. (2020). Testing the efficacy of a

multicomponent, self-guided, smartphone-based meditation app: Three-armed randomized controlled trial. *JMIR Mental Health, 7*, e23825.

Goldberg, S. B., Lam, S. U., Simonsson, O., Torous, J., & Sun, S. (2022). Mobile phone-based interventions for mental health: A systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS Digital Health, 1*, e0000002.

Goldberg, S. B., Riordan, K. M., Sun, S., & Davidson, R. J. (2022). The empirical status of mindfulness-based interventions: A systematic review of 44 meta-analyses of randomized controlled trials. *Perspectives on Psychological Science, 17*, 108-130.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal Of Clinical Psychiatry, 76*, 5356.

Harrer, M., Adam, S. H., Baumeister, H., Cuijpers, P., Karyotaki, E., Auerbach, R. P., . . . Ebert, D. D. (2019). Internet interventions for mental health in university students: A systematic review and meta-analysis. *International Journal of Methods in Psychiatric Research, 28*, e1759.

Heber, E., Ebert, D. D., Lehr, D., Cuijpers, P., Berking, M., Nobis, S., & Riper, H. (2017). The benefit of web-and computer-based interventions for stress: a systematic review and meta-analysis. *Journal Of Medical Internet Research, 19*, e32. doi:10.2196/jmir.5774

Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. United Kingdom, UK: John Wiley & Sons.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558. doi:10.1002/sim.1186

Hirshberg, M. J., Frye, C., Dahl, C. J., Riordan, K. M., Vack, N. J., Sachs, J., . . . Goldberg, S. B. (2022). A Randomized Controlled Trial of a Smartphone-Based Well-Being Training in Public School System Employees During the COVID-19 Pandemic. *Journal of Educational Psychology, 114*, 1895-1911. doi:10.1037/edu0000739

Howells, A., Ivtzan, I., & Eiroa-Orosa, F. J. (2016). Putting the 'app' in happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being, 17*, 163-185. doi:10.1007/s10902-014-9589-1

Huberty, J., Eckert, R., Larkey, L., Kurka, J., De Jesús, S. A. R., Yoo, W., & Mesa, R. (2019). Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. *JMIR Formative Research, 3*, e12662.

Huberty, J., Espel-Huynh, H. M., Neher, T. L., & Puzia, M. E. (2022). Testing the Pragmatic Effectiveness of a Consumer-Based Mindfulness Mobile App in the Workplace: Randomized Controlled Trial. *JMIR mHealth and uHealth, 10*. doi:10.2196/38903

Kabat-Zinn, J. (2007). Full catastrophe living: how to cope with stress, pain and illness using mindfulness meditation ([New edition], reprinted. ed.). *London: Piatkus*.

Keng, S.-L., Chin, J. W. E., Mammadova, M., & Teo, I. (2022). Effects of Mobile App-Based Mindfulness Practice on Healthcare Workers: a Randomized Active Controlled Trial. *Mindfulness, 13*, 2691-2704. doi:10.1007/s12671-022-01975-8

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives Of General Psychiatry, 59*, 877-883. doi:10.1001/archpsyc.59.10.877.

Kranenburg, L. W., Gillis, J., Mayer, B., & Hoogendijk, W. J. G. (2022). The Effectiveness of a Nonguided Mindfulness App on Perceived Stress in a Nonclinical Dutch Population: Randomized Controlled Trial. *JMIR Mental Health, 9*, e32123. doi:10.2196/32123

Kubo, A., Kurtovich, E., McGinnis, M., Aghaee, S., Altschuler, A., Quesenberry, C., . . . Avins, A. L. (2019). A Randomized Controlled Trial of mHealth Mindfulness Intervention for Cancer Patients and Informal Cancer Caregivers: A Feasibility Study Within an Integrated Health Care Delivery System. *Integrative Cancer Therapies, 18*. doi:10.1177/1534735419850634

Kubo, A., Kurtovich, E., McGinnis, M., Aghaee, S., Altschuler, A., Quesenberry, C., . . . Avins, A. (2020). Pilot pragmatic randomized trial of mhealth mindfulness-based intervention for advanced cancer patients and their informal caregivers. *Psycho-Oncology*. doi:10.1002/pon.5557

Lagan, S., D'Mello, R., Vaidyam, A., Bilden, R., & Torous, J. (2021). Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. *Acta Psychiatrica Scandinavica, 144*, 201-210.

Lahtinen, O., Aaltonen, J., Kaakinen, J., Franklin, L., & Hyönä, J. (2021). The effects of app-based mindfulness practice on the well-being of university students and staff. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. doi:10.1007/s12144-021-01762-z

Laird, B., Puzia, M., Larkey, L., Ehlers, D., & Huberty, J. (2022). A Mobile App for Stress Management in Middle-Aged Men and Women (Calm): Feasibility

Randomized Controlled Trial. *JMIR Formative Research, 6*, e30294. doi:10.2196/30294

Lamers, F., van Oppen, P., Comijs, H. C., Smit, J. H., Spinhoven, P., van Balkom, A. J., . . . Penninx, B. W. (2011). Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *The Journal Of Clinical Psychiatry, 72*, 3397.

Lee, R. A., & Jung, M. E. (2018). Evaluation of an mHealth App (DeStressify) on University Students' Mental Health: Pilot Trial. *JMIR mental health, 5*, e2. doi:10.2196/mental.8324

Leng, L. L., Yin, X. C., Chan, C. L. W., & Ng, S. M. (2023). Antenatal mobile-delivered mindfulness-based intervention to reduce perinatal depression risk and improve obstetric and neonatal outcomes: A randomized controlled trial. *Journal Of Affective Disorders, 335*, 216-227. doi:https://doi.org/10.1016/j.jad.2023.04.133

Levin, M. E., Hicks, E. T., & Krafft, J. (2022). Pilot evaluation of the stop, breathe & think mindfulness app for student clients on a college counseling center waitlist. *Journal of American College Health : J of ACH, 70*, 165-173. doi:10.1080/07448481.2020.1728281

Li, A. C. M., Wong, K. K. L., Chio, F. H. N., Mak, W. W. S., & Poon, L. W. H. (2022). Delivering Mindfulness-Based Interventions for Insomnia, Pain, and Dysfunctional Eating Through a Text Messaging App: Three Randomized Controlled Trials Investigating the Effectiveness and Mediating Mechanisms. *Journal of Medical Internet Research, 24*. doi:10.2196/30073

Linardon, J. (2020). Can Acceptance, Mindfulness, and Self-Compassion be learnt by Smartphone Apps? A Systematic and Meta-Analytic Review of

Randomized Controlled Trials. *Behavior Therapy, 51*, 646-658. doi:10.1016/j.beth.2019.10.002

Linardon, J. (2023). Rates of attrition and engagement in randomized controlled trials of mindfulness apps: Systematic review and meta-analysis. *Behaviour Research And Therapy, 170*, 104421. doi:https://doi.org/10.1016/j.brat.2023.104421

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry, 18*, 325-336. doi:10.1002/wps.20673

Linardon, J., & Fuller-Tyszkiewicz, M. (2020). Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal Of Consulting And Clinical Psychology, 88*, 1-13. doi:10.1037/ccp0000459

Linardon, J., Shatte, A., Messer, M., Firth, J., & Fuller-Tyszkiewicz, M. (2020). E-mental health interventions for the treatment and prevention of eating disorders: An updated systematic review and meta-analysis. *Journal Of Consulting And Clinical Psychology, 88*, 994–1007. doi:10.1037/ccp0000575

Linardon, J., Torous, J., Firth, J., Cuijpers, P., Messer, M., & Fuller-Tyszkiewicz, M. (in press). Current evidence for the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. *World Psychiatry*.

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. United Kingdom, UK: Sage Publications.

Lopez, G., Chaoul, A., Warneke, C. L., Christie, A. J., Powers-James, C., Liu, W., . . . Bruera, E. (2023). Self-administered meditation application intervention for cancer patients with psychosocial distress: a pilot study. *Integrative Cancer Therapies, 22*, 15347354221148710.

Luangapichart, P., Saisavoey, N., & Viravan, N. (2022). Efficacy and Feasibility of the Minimal Therapist-Guided Four-Week Online Audio-Based Mindfulness Program 'Mindful Senses' for Burnout and Stress Reduction in Medical Personnel: A Randomized Controlled Trial. *Healthcare, 10*. doi:10.3390/healthcare10122532

Ly, K. H., Trüschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., . . . Andersson, G. (2014). Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *BMJ open, 4*, e003440. doi:10.1136/bmjopen-2013-003440

McClain, L. (2017). *Pilot Randomized Controlled Trial of Smartphone-Delivered Daily Brief Mindfulness Practice for Young Adult Smokers Uninterested in Quitting.* (M.S.), Fordham University, United States -- New York. ProQuest Dissertations & Theses Global database. (10621370)

Miret, M., Ayuso-Mateos, J. L., Sanchez-Moreno, J., & Vieta, E. (2013). Depressive disorders and suicide: epidemiology, risk factors, and burden. *Neuroscience & Biobehavioral Reviews, 37*, 2372-2374.

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. *Journal Of Medical Internet Research, 21*, e12556.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*, 264-269. doi:10.1371/journal.pmed.1000097

Nolan, J. (2020). *Effectiveness, feasibility, and acceptability of a mindfulness based mobile application for undergraduate health science students.* (81), ProQuest Information & Learning, Available from EBSCOhost psyh database.

O'Daffer, A., Colt, S. F., Wasil, A. R., & Lau, N. (2022). Efficacy and conflicts of interest in randomized controlled trials evaluating Headspace and calm apps: systematic review. *JMIR Mental Health, 9*, e40924.

Orosa-Duarte, Á., Mediavilla, R., Muñoz-Sanjose, A., Palao, Á., Garde, J., López-Herrero, V., . . . Rodríguez-Vega, B. (2021). Mindfulness-based mobile app reduces anxiety and increases self-compassion in healthcare students: A randomised controlled trial. *Medical Teacher, 43*, 686-693. doi:10.1080/0142159X.2021.1887835

Pratt, E. H., Hall, L., Jennings, C., Olsen, M. K., Jan, A., Parish, A., . . . Cox, C. E. (2023). Mobile mindfulness for psychological distress and burnout among frontline COVID-19 nurses: a pilot randomized trial. *Annals of the American Thoracic Society, 20*, 1475-1482.

Quinones, C., & Griffiths, M. D. (2019). Reducing compulsive Internet use and anxiety symptoms via two brief interventions: A comparison between mindfulness and gradual muscle relaxation. *Journal of Behavioral Addictions, 8*, 530-536.

Rocamora González, C., Rodríguez Vega, B., Torrijos Zarcero, M., Mediavilla, R., Bouzó Molina, N., Plaza Fernández, R., . . . Palao Tarrero, Á. (2022). Mindfulness based intervention through mobile app for colorectal cancer

people awaiting surgery: A randomized clinical trial. *Cirugia Espanola, 100*, 747-754. doi:10.1016/j.cireng.2022.08.008

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., . . . 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*, 1700-1712.

Segal, Z., Williams, M., & Teasdale, J. (2018). *Mindfulness-based cognitive therapy for depression*: Guilford publications.

Smith, E. N., Santoro, E., Moraveji, N., Susi, M., & Crum, A. J. (2020). Integrating wearables in stress management interventions: Promising evidence from a randomized trial. *International Journal of Stress Management, 27*, 172.

Smith, R. B., Mahnert, N. D., Foote, J., Saunders, K. T., Mourad, J., & Huberty, J. (2021). Mindfulness Effects in Obstetric and Gynecology Patients During the Coronavirus Disease 2019 (COVID-19) Pandemic A Randomized Controlled Trial. *Obstetrics and Gynecology, 137*, 1032-1040. doi:10.1097/aog.0000000000004316

Stoeckl, S. E., Torres-Hernandez, E., Camacho, E., & Torous, J. (2023). Assessing the Dynamics of the Mental Health Apple and Android App Marketplaces. *Journal of Technology in Behavioral Science*, 1-8.

Sun, S. F., Lin, D. H., Goldberg, S., Shen, Z. J., Chen, P. J., Qiao, S., . . . Operario, D. (2022). A Mindfulness-Based Mobile Health (mHealth) Intervention Among Psychologically Distressed University Students in Quarantine During the COVID-19 Pandemic: A Randomized Controlled Trial. *Journal of Counseling Psychology, 69*, 157-171. doi:10.1037/cou0000568

Sun, Y., Li, Y., Wang, J., Chen, Q., Bazzano, A. N., & Cao, F. (2021). Effectiveness of Smartphone-Based Mindfulness Training on Maternal Perinatal Depression: Randomized Controlled Trial. *Journal of Medical Internet Research, 23*, e23410. doi:10.2196/23410

Tan, Z. Y. A., Wong, S. H., Cheng, L. J., & Lau, S. T. (2022). Effectiveness of Mobile-Based Mindfulness Interventions in Improving Mindfulness Skills and Psychological Outcomes for Adults: a Systematic Review and Meta-regression. *Mindfulness, 13*, 2379-2395. doi:10.1007/s12671-022-01962-z

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., . . . Firth, J. (2021). The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry, 20*, 318-335. doi:https://doi.org/10.1002/wps.20883

Torous, J., & Firth, J. (2016). The digital placebo effect: mobile mental health meets clinical psychiatry. *The Lancet Psychiatry, 3*, 100-102. doi:10.1016/S2215-0366(15)00565-9

Treves, I. N., Olson, H. A., Ozernov-Palchik, O., Li, C. E., Wang, K. L., Arechiga, X. M., . . . Gabrieli, J. D. (2023). At-Home use of App-Based Mindfulness for Children: A Randomized Active-Controlled Trial. *Mindfulness*, 1-17.

van Straten, A., Hill, J., Richards, D., & Cuijpers, P. (2015). Stepped care treatment delivery for depression: a systematic review and meta-analysis. *Psychological Medicine, 45*, 231-246.

Versluis, A., Verkuil, B., Spinhoven, P., & Brosschot, J. F. (2018). Effectiveness of a smartphone-based worry-reduction training for stress reduction: A randomized-controlled trial. *Psychology & Health, 33*, 1079-1099. doi:10.1080/08870446.2018.1456660

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., . . . Chen, A. Z. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*, 1545-1602.

Wittchen, H. U. (2002). Generalized anxiety disorder: prevalence, burden, and cost to society. *Depression And Anxiety, 16*, 162-171.

Yoon, S. I., Lee, S. I., Suh, H. W., Chung, S. Y., & Kim, J. W. (2022). Effects of mobile mindfulness training on mental health of employees: A CONSORT-compliant pilot randomized controlled trial. *Medicine, 101*. doi:10.1097/md.0000000000030260

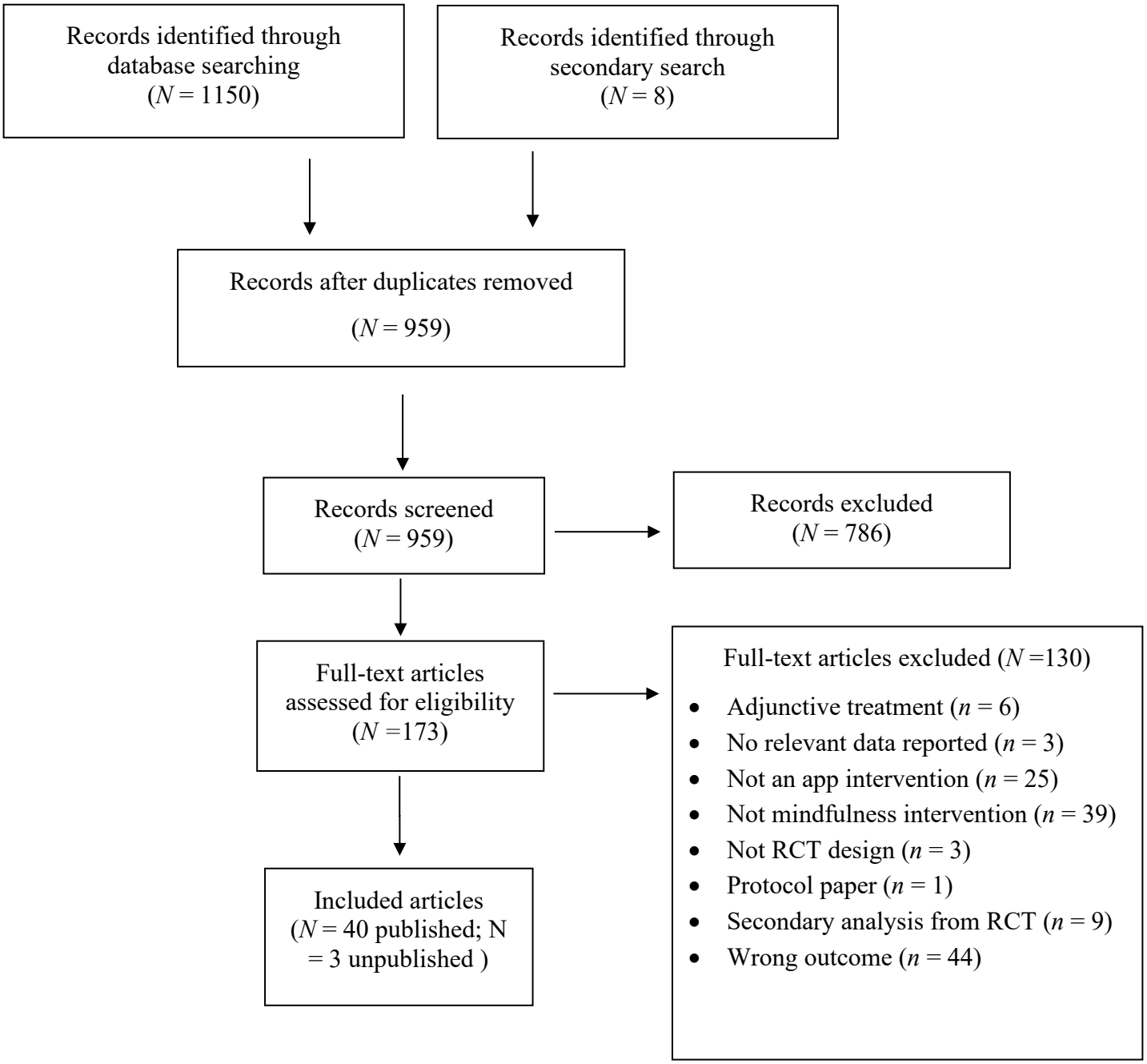

Figure 1: Flowchart of Literature Search

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Abbott, 2018)	Adults with elevated anxiety or worry symptoms	Headspace (n = 97)	Waitlist (n = 66)	4 weeks	No (Fully remote)	No	10-40 min at least 6 days per week	ITT	-	**BAI**
(Bear et al., 2022)	Mothers of children aged 0-12 months	Smiling Mind (n = 49)	Baby + Tracker control app (n = 50)	8 weeks	No (Fully remote)	No	At least one session per day	C	DASS	DASS
(Bhayee et al., 2016)	Adults under self-reported moderate-high stress	Calm (n = 20)	Online math training (n = 20)	6 weeks	Yes (In-person)	Yes	Minimum of 32/42 sessions over 6 weeks	C	BSI	BSI
(Borjalilu et al., 2019)	Students with elevated stress	Armagar (n = 20)	Face-to-face mindfulness (n = 20)	3 weeks	Yes (In-person)	No	-	ITT	DASS	DASS
(Boden et al., 2023)	Orthopaedic surgery residents	Headspace (n = 12)	Waitlist (n = 12)	8 weeks	No (Fully remote)	No	-	ITT		**GAD-7**
(Bosso, 2020)	Students	Headspace (n = 22)	Waitlist (n = 22)	5 week	Yes (In-person)	No	10 minutes daily	ITT	DASS	DASS
(Bostock et al., 2019)	Employees with elevated stress	Headspace (n = 128)	Educational material on work stress (n = 120)	8 weeks	Yes (In-person)	No	One session per day	C	HADS	HADS

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Carissoli et al., 2015)	Adult employees	It's time to relax (n = 20)	Music listening control (n = 18)	3 weeks	Yes (In-person)	No	Two, 15 minute meditation sessions per day	ITT	MSP	MSP
(Cox et al., 2019)	Adult ICU patients	Mindfulness app (n = 31)	Telephone mindfulness (n = 31) Education control (n = 18)	4 weeks	Yes (In-person)	Yes	-	C	PHQ-9	GAD-7
(Fish & Saul, 2019)	Students	Headspace (n = 47)	Waitlist (n = 44)	2 weeks	No (Fully remote)	No	One mindfulness session per day	C	**PHQ-9**	
(Flett et al., 2019)	Students	Headspace (n = 72) Smiling Mind (n = 63)	Evernote control app (n = 75)	4 weeks	Yes (In-person)	No	10 minutes each day	C	**CES-D**	**HADS**
(Forbes et al., 2020)	Women with chronic pelvis pain	Headspace (n = 31)	Usual care (n = 29) Muscle relaxation control (n = 30)	8 weeks	Yes (in person)	No	-	C	HADS	HADS

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Gao et al., 2022)	Adults with anxiety and sleep disturbances	Unwinding anxiety (n = 40)	Usual care (n = 40)	8 weeks	Yes (In-person)	Yes		C		GAD-7
(Goldberg et al., 2020)	General population	Connection (n = 121) Insight (n = 107)	Waitlist (n = 115)	8 weeks	No (Fully remote)	Yes	-	C	PROMIS	PROMIS
(Hirshberg et al., 2022)	Employees	Healthy Minds (n = 346)	Waitlist (n = 320)	4 weeks	No (Fully remote)	Yes	-	C	PROMIS	PROMIS
(Howells et al., 2016)	General population of adults	Headspace (n = 97)	Catch notes control app (n = 97)	10 days	No (Fully remote)	No	10 minutes each day	C	CES-D	
(Huberty et al., 2019)	Myeloproliferative neoplasm patients	10% Happier (n = 33) Calm (n = 32)	Educational control (n = 63)	5 weeks	No (Fully remote)	No	-	C	PROMIS	PROMIS
(Huberty et al., 2022)	Employees	Calm (n = 585)	Waitlist (n = 444)	8 weeks	No (Fully remote)	No	10 minutes each day	C	DASS	DASS
(Keng et al., 2022)	Healthcare workers	Headspace (n = 40)	Lumosity app control (n = 40)	3 weeks	Yes (Telephone)	Yes	10 minutes each day	C	DASS	DASS

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Kranenburg et al., 2022)	General population of adults	Mindfulness app (n = 386)	Educational resources (n = 425	8 weeks	No (Fully remote)	No	-	C	4DSQ	4DSQ
(Kubo et al., 2019)	Cancer patients	Headspace (n = 54)	Waitlist (n = 43)	8 weeks	Yes (Telephone)	Yes	10-20 minutes each day	C	HADS	HADS
(Kubo et al., 2020)	Cancer patients	Headspace (n = 52)	Waitlist (n = 51)	6 weeks	Yes (Telephone)	Yes	-	C	HADS	HADS
(Lahtinen et al., 2021)	University staff and students	Welzen (n = 282)	Psychoeducational control (n = 279)	4 weeks	No (Fully remote)	No	10 minutes each day	C	**BDI**	**GAD-7**
(Laird et al., 2022)	Adults with elevated stress	Calm (n = 39)	Psychoeducational control (n = 35)	4 weeks	No (Fully remote)	Yes	10 minutes each day	C	HADS	HADS
(Lee & Jung, 2018)	Students	DeStressify (n = 102)	Waitlist (n = 104)	4 weeks	No (Fully remote)	Yes	5 days per week over 4 weeks	C	QIDS-SR	STAI
(Leng et al., 2023)	Women with elevated stress	Thrive Pregnancy (n = 38)	Education control (n = 37)	8 weeks	Yes (telephone)	No	-	ITT	EPDS	
(Levin et al., 2022)	University students	Stop, Breathe, Think (n = 10)	Waitlist (n = 12)	4 weeks	No (Fully remote)	No	-	C	CCAPS	CCAPS
(Li et al., 2022) – Study 1	General population adults	WhatsApp (n = 167)	Waitlist (n = 166)	22 days	No (Fully remote)	No	-	ITT	PHQ-9	GAD-7

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Li et al., 2022) – Study 2	General population adults	WhatsApp (n = 118)	Waitlist (n = 117)	22 days	No (Fully remote)	No	-	ITT	PHQ-9	GAD-7
(Li et al., 2022) – Study 3	General population adults	WhatsApp (n = 177)	Waitlist (n = 174)	22 days	No (Fully remote)	No	-	ITT	PHQ-9	GAD-7
(Lopez et al., 2023)	Cancer patients with elevated distress	Mindfulness app (n = 17)	Waitlist (n = 18)	2 weeks	Yes (in person)	No	-	C	ESAS-FS	ESAS-FS
(Luangapichart et al., 2022)	Healthcare workers with elevated stress	Mindful Senses (n = 45)	Education control (n = 45)	4 weeks	No (Fully remote)	Yes	Engage in meditation at least 3 times per day	ITT	HADS	HADS
(Ly et al., 2014)	Adults with elevated depression	Mindfulness app (n = 41)	Behavioral activation app (n = 40)	8 weeks	Yes (Telephone)	No	-	ITT	**BDI PHQ-9**	BAI
(Nolan, 2020)	Students	Headspace (n = 49)	Waitlist (n = 46)	10 days	Yes (In-person)	Yes	10 minutes each day	C	DASS	DASS
(Orosa-Duarte et al., 2021)	Healthcare students	Rem Volver a Casa' app (n = 54)	Face-to-Face mindfulness therapy (n = 51) Waitlist (n = 49)	8 weeks	?	No	1 stage per week	C		**STAI**

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	Anx
(Pratt et al., 2023)	Nurses	Life (n = 33)	Waitlist (n = 33)	4 weeks	No (Fully remote)	No	-	C	PHQ-9	GAD-7
(Quinones & Griffiths, 2019)	Compulsive internet users	Headspace (n = 343)	Waitlist (n = 350) Muscle relaxation control (n= 301)	2 weeks	No (Fully remote)	No	10 minutes each day	C	PHQ-2	PHQ-2
(Rocamora González et al., 2022)	Patients with colorectal cancer	Calm in the Operating Room (n = 52)	Care as usual (n = 50)	?	?	No	-	C	HADS	HADS
(Treves et al., 2023)	Children	Inner Explorer (n = 101)	Audiobook control (n = 105) Audiobook + scaffolder (n = 108)	8 weeks	Yes (Zoom)	No	10 minutes each day	C	RCADS-25-C	**RCADS-25-C**
(E. N. Smith et al., 2020)	Employees	Spire (n = 107)	Waitlist (n = 108)	4 weeks	No (fully remote)	No	6-9 min session per week	C	MASQ	**MASQ**
(R. B. Smith et al., 2021)	Obstetric patients	Calm (n = 50)	Usual care (n = 51)	4 weeks	Yes (Telephone)	No	10 minutes each day	ITT	HADS	HADS
(S. F. Sun et al., 2022)	University students with elevated distress	Mindfulness for Growth and Resilience (n = 57)	Social support app (n = 57)	4 weeks	Yes (Video-conference)	No	5-10 minutes each day	ITT	**PHQ-9**	**GAD-7**

Table 1
Characteristics of Included Randomized Controlled Trials

Study	Sample characteristics	Mindfulness app (N randomized)	Comparison group (*N* randomized)	Post-test length	Researcher contact?	Monetary comp?	Recommended practice	Analysis used for ES calc	Outcome	
									Dep	**Anx**
(Y. Sun et al., 2021)	Pregnant women with depressive symptoms	Spirits Healing (n = 84)	Mobile health consultations control (n = 84)	8 weeks	Yes (Telephone)	Yes	Daily practice	ITT	**EPDS**	GAD-7
(Versluis et al., 2018)	Adults with elevated stress	VGZ Mindfulness Coach (n = 46)	Waitlist (n = 48) Placebo control app (n = 42)	4 weeks	Yes (Telephone)	No	-	C	PHQ-9	GAD-7
(Yoon et al., 2022)	Employees with elevated stress	InMind app (n = 22)	Waitlist (n = 23)	4 weeks	?	Yes	Daily practice	ITT	MBI	MBI

C = complete case; ITT = intention-to-treat; analysis = data used for effect size calculation; MBI = Mibyeong Index; PHQ = Patient Health Questionnaire; GAD = Generalized Anxiety Disorder Scale; EPDS = Edinburgh Postnatal Depression Scale; HADS = Hospital for Depression Anxiety Scale; DASS = Depression Anxiety and Stress Scale; STAI = State Trait Anxiety Scale; BAI = Beck Anxiety Inventory; CCAPS = Counseling Center Assessment of Psychological Symptoms; QUIDS = The Quick Inventory of Depressive Symptomatology; CES-D = Centre for Epidemiology Studies – Depression; BSI = Brief Symptom Inventory; MSP = Mesure du Stress Psychologique; PROMIS = Reported Outcomes Measurement Information System Inventory. Bold denotes that the measure of depression or anxiety was declared as the primary outcome.

Table 2
Results from the Meta-Analyses and Subgroup Analyses

	Depressive symptoms				Anxiety symptoms			
Analysis	N_{comp}	***g* (95% CI)**	I^2	***p***	N_{comp}	***g* (95% CI)**	I^2	***p***
Mindfulness apps vs. control conditions								
Total effect	46	0.24 (0.17, 0.31)	32%		48	0.28 (0.21, 0.35)	44%	
Trim and Fill	46	0.24 (0.17, 0.31)	32%		42	0.33 (0.25, 0.41)		
One effect per study (smallest)	37	0.26 (0.18, 0.34)	37%		42	0.28 (0.20,0.36)	48%	
One effect per study (largest)	38	0.27 (0.20, 0.34)	31%		42	0.29 (0.21, 0.38)	46%	
Low risk of bias only	7	0.29 (0.16, 0.42)	18%		9	0.42 (0.22, 0.61)	62%	
Larger studies only (> 99 per condition)	7	0.31 (0.17, 0.46)	65%		7	0.33 (0.24, 0.41)	14%	
Subgroup analysis								
Control condition				.129				.213
Inactive	21	0.30 (0.21, 0.38)	36%		25	0.32 (0.24, 0.41)	28%	
Active	25	0.20 (0.10, 0.30)	16%		23	0.23 (0.11, 0.35)	53%	
Assessment length				.242				.846
≤ 4 weeks	27	0.28 (0.20, 0.36)	28%		26	0.30 (0.22, 0.38)	27%	
> 4 weeks	18	0.19 (0.08, 0.31)	30%		21	0.28 (0.15, 0.41)	48%	
Sample type				.438				.296
Pre-existing mental health problems	11	0.30 (0.12, 0.47)	44%		12	0.36 (0.18, 0.53)	51%	
No pre-existing mental health problems	35	0.22 (0.15, 0.30)	28%		36	0.25 (0.17, 0.33)	42%	
Researcher contact				.336				.294
Yes (in-person/telephone)	24	0.21 (0.10, 0.32)	30%		24	0.24 (0.14, 0.33)	14%	
No (fully remote)	20	0.28 (0.19, 0.36)	29%		21	0.31 (0.22, 0.40)	43%	
Monetary compensation				**.046**				.203
Yes	15	0.34 (0.25, 0.44)	0%		16	0.35 (0.20, 0.50)	51%	
No	31	0.21 (0.13, 0.30)	40%		32	0.24 (0.16, 0.32)	36%	
Headspace app				.373				.992

Yes	14	0.28 (0.18, 0.39)	0%		14	0.28 (0.15, 0.42)	28%	
No	32	0.22 (0.14, 0.31)	45%		34	0.28 (0.19, 0.37)	50%	
Calm app				.778				.374
Yes	5	0.20 (-0.05, 0.46)	46%		5	0.22 (0.04, 0.40)	10%	
No	41	0.24 (0.17, 0.31)	32%		43	0.28 (0.20, 0.36)	47%	
Mindfulness apps vs. active comparisons								
Total effect	3	-0.15 (-0.44, 0.12)	0%		4	0.10 (-0.29, 0.51)	58%	
Trim and fill method	3	-0.15 (-0.44, 0.12)			3	0.18 (-0.17, 0.53)		
Low risk of bias only	2	-0.13 (-0.46, 0.20)	0%		2	-0.07 (-0.41, 0.25)	0%	

Supplementary Materials

Table 2
Risk of Bias Domain Ratings Across Studies

Study	Random sequence	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Total
(Abbott, 2018)	Low	Unclear	High	Low	Low	3
(Bear et al., 2022)	Unclear	Unclear	Low	Low	High	2
(Bhayee et al., 2016)	Low	Low	Low	Low	Low	5
(Borjalilu et al., 2019)	Unclear	Unclear	Low	Low	Low	3
(Bosso, 2020)	Unclear	Unclear	Low	Low	Low	3
(Bostock et al., 2019)	Low	Unclear	High	Low	High	2
(Boden et al., 2023)	Unclear	Unclear	High	Low	Low	2
(Carissoli et al., 2015)	Unclear	Unclear	Low	Low	Low	3
(Cox et al., 2019)	Low	Unclear	Low	Low	Low	4
(Fish & Saul, 2019)	Unclear	Unclear	High	Low	High	1
(Flett et al., 2019)	Low	Unclear	Low	Low	High	3
(Forbes et al., 2020)	Low	Low	High	Low	High	3
(Gao et al., 2022)	Low	Low	High	Low	Low	4
(Goldberg et al., 2020)	Low	Unclear	High	Low	Low	3
(Hirshberg et al., 2022)	Low	Unclear	High	Low	Low	3
(Howells et al., 2016)	Unclear	Unclear	Low	Low	High	2
(Huberty et al., 2019)	Low	Unclear	Low	Low	High	3
(Huberty et al., 2022)	Low	Low	High	Low	Low	4
(Keng et al., 2022)	Low	Unclear	Low	Low	High	3
(Kubo et al., 2019)	Low	Low	High	Low	High	3
(Kubo et al., 2020)	Unclear	Unclear	High	Low	High	1
(Kranenburg et al., 2022)	Low	Unclear	High	Low	High	2
(Lahtinen et al., 2021)	Low	Unclear	High	Low	Low	4
(Laird et al., 2022)	Low	Unclear	Low	Low	High	3
(Lee & Jung, 2018)	Low	Unclear	High	Low	High	2
(Levin et al., 2022)	Low	Unclear	High	Low	High	2
(Li et al., 2022) – Study 1	Unclear	Unclear	High	Low	Low	2
(Li et al., 2022) – Study 2	Unclear	Unclear	High	Low	Low	2
(Li et al., 2022) – Study 3	Unclear	Unclear	High	Low	Low	2

(Lopez et al., 2023)	Low	Unclear	High	Low	High	2
(Luangapichart et al., 2022)	Low	Unclear	Low	Low	Low	4
(Ly et al., 2014)	Low	Low	Low	Low	Low	5
(Nolan, 2020)	Unclear	Unclear	High	Low	High	1
(Orosa-Duarte et al., 2021)	Low	Unclear	High	Low	High	2
(Pratt et al., 2023)	Low	Unclear	High	Low	High	2
(Quinones & Griffiths, 2019)	Unclear	Unclear	High	Low	High	1
(Rocamora González et al., 2022)	Low	Unclear	High	Low	Low	3
(Treves et al., 2023)	Unclear	Unclear	Low	Low	High	2
(R. B. Smith et al., 2021)	Low	Low	High	Low	Low	4
(E. N. Smith et al., 2020)	Unclear	Unclear	High	Low	High	1
(S. F. Sun et al., 2022)	Low	Low	Low	Low	Low	5
(Y. Sun et al., 2021)	Low	Low	Low	Low	Low	5
(Versluis et al., 2018)	Low	Low	High	Low	High	3
(Yoon et al., 2022)	Low	Unclear	High	Low	Low	3

Table S1 – high level summary of differences between within group (intervention) and between group effects for anxiety symptoms

Study	$SMD_{withinINT}$ (low, high)	$SMD_{between}$
Abbott 2018	0.28 (0.18, 0.54)	0.14
Bear et al 2022	0.53 (0.38, 0.82)	0.57
Bhayee et al 2016	0.48 (0.38, 0.65)	0.29
Borjalilu et al 2019	0.57 (0.37, 1.05)	-0.14
Bosso, 2020	0.83 (0.60, 1.29)	0.32
Bostock et al 2019	0.52 (0.34, 0.99)	0.38
Carissoli et al 2015	-0.06 (-0.04, -0.11)	0.13
Flett et al 2019a	0.24 (0.16, 0.47)	0.15

Flett et al 2019b	0.30 (0.19, 0.56)	0.12
Forbes 2020a	0.02 (0.01, 0.04)	-0.61
Forbes 2020b	0.02 (0.01, 0.04)	-0.36
Gao et al 2022	0.90 (0.58, 1.64)	0.71
Huberty et al 2022	0.55 (0.35, 1.03)	0.28
Keng et al 2022	0.27 (0.18, 0.48)	0.34
Kubo et al 2019	-0.04 (-0.02, -0.07)	0.09
Kubo et al 2019	0.42 (0.27, 0.80)	-0.16
Laird et al., 2022	-0.03 (-0.02, -0.05)	-0.17
Lee & Jung, 2018	0.33 (0.22, 0.63)	0.30
Levin et al., 2022	0.94 (0.70, 1.36)	0.64
Luangapichart et al., 2022	1.44 (0.94, 2.64)	1.13
Ly et al., 2014	0.57 (0.37, 1.04)	-0.06
Nolan, 2020	0.28 (0.19, 0.50)	-0.04
Orosa-Duarte et al 2021	1.52 (1.24, 1.93)	0.77
Orosa-Duarte et al 2021	1.52 (1.24, 1.93)	1.15
Pratt et al. 2023	0.63 (0.40, 1.20)	0.35
Rocamora González et al., 2022	0.17 (0.11, 0.31)	-0.49
Smith et al 2021	0.73 (0.47, 1.39)	0.44
Sun et al 2021	0.41 (0.27, 0.76)	0.48
Sun et al 2022	1.61 (1.05, 2.97)	0.36
Treves 2023a	0.19 (0.12, 0.36)	0.03
Treves 2023b	0.21 (0.13, 0.40)	-0.20
Treves 2023c	0.19 (0.12, 0.36)	0.01
Treves 2023d	0.21 (0.13, 0.40)	0.03
Versluis et al 2018a	0.23 (0.15, 0.44)	0.05
Versluis et al 2018b	0.23 (0.15, 0.44)	0.28
Yoon et al 2022	0.57 (0.36, 1.08)	0.42

Notes. Shading indicates between group effect is larger (more positive) than within group effect. Low = effect when r = .1, high = effect when r = .9

Table S2 – high level summary of differences between within group (intervention) and between group effects for depressive symptoms

Study	$SMD_{withinINT}$ (low, high)	$SMD_{between}$

Bear et al 2022	0.77 (0.50, 1.44)	0.51
Bhayee et al 2016	0.21 (0.14, 0.36)	-0.03
Borjalilu et al 2019	0.66 (0.43, 1.21)	-0.25
Bosso, 2020	0.70 (0.56, 0.92)	-0.23
Bostock et al 2019	0.51 (0.33, 0.97)	0.47
Carissoli et al 2015	-0.06 (-0.04, -0.11)	0.13
Flett et al 2019a	0.22 (0.14, 0.43)	0.17
Flett et al 2019b	0.49 (0.35, 0.79)	0.12
Forbes 2020a	0.36 (0.23, 0.69)	0.28
Forbes 2020b	0.36 (0.23, 0.69)	0.25
Howells et al 2016	0.44 (0.29, 0.78)	0.36
Huberty et al 2022	0.69 (0.45, 1.28)	0.41
Keng et al 2022	0.31 (0.22, 0.51)	0.44
Kubo et al 2019	0.33 (0.22, 0.58)	0.50
Kubo et al 2019	0.36 (0.23, 0.69)	0.09
Laird et al., 2022	-0.24 (-0.15, -0.45)	-0.22
Lee & Jung, 2018	0.57 (0.36, 1.07)	0.23
Leng et al. 2023	1.16 (0.83, 1.82)	0.87
Levin et al., 2022	1.43 (0.96, 2.47)	0.48
Luangapichart et al., 2022	1.77 (1.13, 3.38)	0.44
Ly et al., 2014	1.39 (0.89, 2.64)	-0.25
Nolan, 2020	0.27 (0.17, 0.50)	0.23
Pratt et al. 2023	0.82 (0.53, 1.54)	0.52
Rocamora González et al., 2022	-0.34 (-0.22, -0.65)	-0.22
Smith et al 2021	0.46 (0.30, 0.84)	0.44
Sun et al 2021	0.51 (0.33, 0.96)	0.27
Sun et al 2022	1.69 (1.09, 3.12)	0.41
Treves 2023a	0.18 (0.12, 0.34)	-0.04
Treves 2023b	0.20 (0.13, 0.36)	-0.02
Treves 2023c	0.18 (0.12, 0.34)	-0.14
Treves 2023d	0.20 (0.13, 0.36)	0.03
Versluis et al 2018a	0.34 (0.30, 0.40)	-0.12

Versluis et al 2018b	0.34 (0.30, 0.40)	0.08
Yoon et al 2022	0.56 (0.37, 0.99)	0.50

Notes. Shading indicates between group effect is larger (more positive) than within group effect. Low = effect when r = .1, high = effect when r = .9

Table S3 – anxiety as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .1 to .3

	r = .1			*r* = .2			*r* = .3		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Abbott 2018	0.18	-0.08	0.14	0.19	-0.09	0.14	0.21	-0.09	0.14
Bear et al 2022	0.38	0.01	0.57	0.40	0.01	0.57	0.43	0.01	0.57
Bhayee et al 2016	0.38	0.03	0.29	0.39	0.03	0.29	0.41	0.03	0.29
Boden 2023	0.94	0.14	-0.48	1.00	0.14	-0.48	1.06	0.15	-0.48
Borjalilu et al 2019	0.37	0.43	-0.14	0.39	0.46	-0.14	0.42	0.49	-0.14
Bosso, 2020	0.60	0.36	0.32	0.63	0.38	0.32	0.67	0.41	0.32
Bostock et al 2019	0.34	0.09	0.38	0.36	0.10	0.38	0.38	0.11	0.38
Carissoli et al 2015	-0.04	-0.06	0.13	-0.04	-0.06	0.13	-0.05	-0.07	0.13
Flett et al 2019a	0.16	-0.03	0.15	0.17	-0.04	0.15	0.18	-0.04	0.15
Flett et al 2019b	0.19	-0.03	0.12	0.20	-0.04	0.12	0.21	-0.04	0.12
Forbes 2020a	0.01	0.39	-0.61	0.01	0.42	-0.61	0.02	0.44	-0.61
Forbes 2020b	0.01	0.04	-0.36	0.01	0.04	-0.36	0.02	0.04	-0.36
Gao et al 2022	0.58	-0.13	0.71	0.62	-0.14	0.71	0.66	-0.15	0.71
Huberty et al 2022	0.35	0.21	0.28	0.38	0.23	0.28	0.40	0.24	0.28
Keng et al 2022	0.18	0.07	0.34	0.19	0.07	0.34	0.20	0.08	0.34
Kubo et al 2019	-0.02	0.04	0.09	-0.03	0.05	0.09	-0.03	0.05	0.09
Kubo et al 2019	0.27	0.33	-0.16	0.29	0.35	-0.16	0.31	0.37	-0.16
Laird et al., 2022	-0.02	-0.23	-0.17	-0.02	-0.25	-0.17	-0.02	-0.27	-0.17
Lee & Jung, 2018	0.22	0.01	0.30	0.23	0.01	0.30	0.24	0.01	0.30
Levin et al., 2022	0.70	0.17	0.64	0.74	0.18	0.64	0.78	0.19	0.64
Luangapichart et al., 2022	0.94	0.45	1.13	1.00	0.48	1.13	1.06	0.51	1.13

Ly et al., 2014	0.37	0.56	-0.06	0.40	0.59	-0.06	0.42	0.63	-0.06
Nolan, 2020	0.19	0.20	-0.04	0.20	0.21	-0.04	0.21	0.22	-0.04
Orosa-Duarte et al 2021	1.24	0.21	0.77	1.29	0.22	0.77	1.34	0.24	0.77
Orosa-Duarte et al 2021	1.24	-0.22	1.15	1.29	-0.22	1.15	1.34	-0.23	1.15
Pratt et al., 2023	0.40	0.17	0.35	0.43	0.18	0.35	0.45	0.19	0.35
Rocamora González et al., 2022	0.11	0.09	-0.49	0.12	0.09	-0.49	0.13	0.10	-0.49
Smith et al 2021	0.47	0.23	0.44	0.50	0.24	0.44	0.53	0.26	0.44
Sun et al 2021	0.27	-0.07	0.48	0.28	-0.08	0.48	0.30	-0.08	0.48
Sun et al 2022	1.05	0.51	0.36	1.11	0.54	0.36	1.19	0.57	0.36
Treves 2023a	0.12	0.05	0.03	0.13	0.05	0.03	0.14	0.05	0.03
Treves 2023b	0.13	0.06	-0.20	0.14	0.07	-0.20	0.15	0.07	-0.20
Treves 2023c	0.12	0.06	0.01	0.13	0.07	0.01	0.14	0.07	0.01
Treves 2023d	0.13	0.01	0.03	0.14	0.01	0.03	0.15	0.01	0.03
Versluis et al 2018a	0.15	0.03	0.05	0.16	0.03	0.05	0.17	0.03	0.05
Versluis et al 2018b	0.15	-0.08	0.28	0.16	-0.08	0.28	0.17	-0.09	0.28
Yoon et al 2022	0.36	-0.26	0.42	0.38	-0.28	0.42	0.41	-0.30	0.42

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores

Table S4 – anxiety as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .4 to .6

	r = .4			*r* = .5			*r* = .6		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Abbott 2018	0.22	-0.10	0.14	0.24	-0.11	0.14	0.27	-0.12	0.14
Bear et al 2022	0.46	0.01	0.57	0.49	0.01	0.57	0.54	0.01	0.57
Bhayee et al 2016	0.44	0.04	0.29	0.46	0.04	0.29	0.49	0.04	0.29
Boden 2023	1.13	0.17	-0.48	1.23	0.18	-0.48	1.35	0.20	-0.48
Borjalilu et al 2019	0.45	0.53	-0.14	0.49	0.58	-0.14	0.55	0.65	-0.14
Bosso, 2020	0.71	0.44	0.32	0.77	0.48	0.32	0.84	0.53	0.32
Bostock et al 2019	0.41	0.12	0.38	0.45	0.13	0.38	0.50	0.14	0.38
Carissoli et al 2015	-0.05	-0.07	0.13	-0.06	-0.08	0.13	-0.06	-0.09	0.13

Flett et al 2019a	0.19	-0.04	0.15	0.21	-0.05	0.15	0.23	-0.05	0.15
Flett et al 2019b	0.23	-0.04	0.12	0.25	-0.05	0.12	0.28	-0.05	0.12
Forbes 2020a	0.02	0.48	-0.61	0.02	0.52	-0.61	0.02	0.58	-0.61
Forbes 2020b	0.02	0.05	-0.36	0.02	0.05	-0.36	0.02	0.06	-0.36
Gao et al 2022	0.71	-0.16	0.71	0.78	-0.18	0.71	0.87	-0.20	0.71
Huberty et al 2022	0.43	0.26	0.28	0.47	0.29	0.28	0.53	0.32	0.28
Keng et al 2022	0.22	0.09	0.34	0.24	0.09	0.34	0.26	0.10	0.34
Kubo et al 2019	-0.03	0.05	0.09	-0.03	0.06	0.09	-0.04	0.07	0.09
Kubo et al 2019	0.33	0.40	-0.16	0.36	0.44	-0.16	0.41	0.49	-0.16
Laird et al., 2022	-0.02	-0.29	-0.17	-0.02	-0.31	-0.17	-0.03	-0.35	-0.17
Lee & Jung, 2018	0.26	0.01	0.30	0.29	0.01	0.30	0.32	0.01	0.30
Levin et al., 2022	0.82	0.21	0.64	0.88	0.23	0.64	0.95	0.25	0.64
Luangapichart et al., 2022	1.15	0.55	1.13	1.25	0.60	1.13	1.39	0.67	1.13
Ly et al., 2014	0.46	0.67	-0.06	0.50	0.73	-0.06	0.55	0.80	-0.06
Nolan, 2020	0.23	0.24	-0.04	0.25	0.26	-0.04	0.28	0.29	-0.04
Orosa-Duarte et al 2021	1.40	0.25	0.77	1.47	0.27	0.77	1.56	0.29	0.77
Orosa-Duarte et al 2021	1.40	-0.24	1.15	1.47	-0.25	1.15	1.56	-0.26	1.15
Pratt et al. 2023	0.49	0.21	0.35	0.54	0.22	0.35	0.60	0.25	0.35
Rocamora González et al., 2022	0.14	0.11	-0.49	0.15	0.12	-0.49	0.17	0.13	-0.49
Smith et al 2021	0.58	0.28	0.44	0.63	0.31	0.44	0.71	0.34	0.44
Sun et al 2021	0.33	-0.09	0.48	0.36	-0.10	0.48	0.40	-0.11	0.48
Sun et al 2022	1.28	0.62	0.36	1.40	0.67	0.36	1.56	0.75	0.36
Treves 2023a	0.15	0.06	0.03	0.16	0.06	0.03	0.18	0.07	0.03
Treves 2023b	0.16	0.08	-0.20	0.18	0.08	-0.20	0.20	0.09	-0.20
Treves 2023c	0.15	0.08	0.01	0.16	0.08	0.01	0.18	0.09	0.01
Treves 2023d	0.16	0.01	0.03	0.18	0.01	0.03	0.20	0.01	0.03
Versluis et al 2018a	0.18	0.04	0.05	0.20	0.04	0.05	0.22	0.04	0.05
Versluis et al 2018b	0.18	-0.10	0.28	0.20	-0.11	0.28	0.22	-0.12	0.28
Yoon et al 2022	0.44	-0.32	0.42	0.49	-0.35	0.42	0.54	-0.39	0.42

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores

Table S5 – anxiety as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .7 to .9

	r = .7			*r* = .8			*r* = .9		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Abbott 2018	0.31	-0.14	0.14	0.38	-0.17	0.14	0.54	-0.24	0.14
Bear et al 2022	0.60	0.01	0.57	0.68	0.02	0.57	0.82	0.02	0.57
Bhayee et al 2016	0.53	0.05	0.29	0.58	0.06	0.29	0.65	0.09	0.29
Boden 2023	1.52	0.23	-0.48	1.78	0.28	-0.48	2.24	0.40	-0.48
Borjalilu et al 2019	0.63	0.75	-0.14	0.76	0.92	-0.14	1.05	1.29	-0.14
Bosso, 2020	0.93	0.61	0.32	1.07	0.73	0.32	1.29	0.98	0.32
Bostock et al 2019	0.58	0.16	0.38	0.71	0.20	0.38	0.99	0.28	0.38
Carissoli et al 2015	-0.07	-0.10	0.13	-0.08	-0.12	0.13	-0.11	-0.17	0.13
Flett et al 2019a	0.27	-0.06	0.15	0.33	-0.07	0.15	0.47	-0.09	0.15
Flett et al 2019b	0.33	-0.06	0.12	0.40	-0.07	0.12	0.56	-0.09	0.12
Forbes 2020a	0.02	0.66	-0.61	0.03	0.79	-0.61	0.04	1.04	-0.61
Forbes 2020b	0.02	0.06	-0.36	0.03	0.08	-0.36	0.04	0.11	-0.36
Gao et al 2022	0.99	-0.23	0.71	1.20	-0.28	0.71	1.64	-0.39	0.71
Huberty et al 2022	0.61	0.37	0.28	0.74	0.45	0.28	1.03	0.64	0.28
Keng et al 2022	0.30	0.12	0.34	0.36	0.15	0.34	0.48	0.21	0.34
Kubo et al 2019	-0.04	0.08	0.09	-0.05	0.09	0.09	-0.07	0.13	0.09
Kubo et al 2019	0.47	0.56	-0.16	0.57	0.69	-0.16	0.80	0.97	-0.16
Laird et al., 2022	-0.03	-0.40	-0.17	-0.04	-0.48	-0.17	-0.05	-0.64	-0.17
Lee & Jung, 2018	0.37	0.01	0.30	0.45	0.01	0.30	0.63	0.02	0.30
Levin et al., 2022	1.05	0.29	0.64	1.17	0.36	0.64	1.36	0.50	0.64
Luangapichart et al., 2022	1.60	0.77	1.13	1.93	0.93	1.13	2.64	1.28	1.13
Ly et al., 2014	0.64	0.89	-0.06	0.77	1.02	-0.06	1.04	1.25	-0.06
Nolan, 2020	0.32	0.33	-0.04	0.38	0.40	-0.04	0.50	0.54	-0.04
Orosa-Duarte et al 2021	1.66	0.31	0.77	1.78	0.34	0.77	1.93	0.39	0.77
Orosa-Duarte et al 2021	1.66	-0.28	1.15	1.78	-0.29	1.15	1.93	-0.31	1.15
Pratt et al. 2023	0.69	0.28	0.35	0.85	0.33	0.35	1.20	0.41	0.35
Rocamora González et al., 2022	0.19	0.15	-0.49	0.23	0.19	-0.49	0.31	0.26	-0.49
Smith et al 2021	0.81	0.40	0.44	0.99	0.48	0.44	1.39	0.68	0.44
Sun et al 2021	0.46	-0.12	0.48	0.55	-0.14	0.48	0.76	-0.19	0.48
Sun et al 2022	1.79	0.87	0.36	2.17	1.05	0.36	2.97	1.45	0.36
Treves 2023a	0.21	0.08	0.03	0.26	0.10	0.03	0.36	0.14	0.03

Treves 2023b	0.23	0.11	-0.20	0.29	0.13	-0.20	0.40	0.18	-0.20
Treves 2023c	0.21	0.11	0.01	0.26	0.13	0.01	0.36	0.18	0.01
Treves 2023d	0.23	0.01	0.03	0.29	0.01	0.03	0.40	0.02	0.03
Versluis et al 2018a	0.25	0.05	0.05	0.31	0.06	0.05	0.44	0.08	0.05
Versluis et al 2018b	0.25	-0.14	0.28	0.31	-0.17	0.28	0.44	-0.23	0.28
Yoon et al 2022	0.63	-0.44	0.42	0.77	-0.53	0.42	1.08	-0.71	0.42

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores

Table S6– depression as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .1 to .3

	r = .1			*r* = .2			*r* = .3		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Bear et al 2022	0.50	0.14	0.51	0.53	0.15	0.51	0.56	0.16	0.51
Bhayee et al 2016	0.14	0.00	-0.03	0.15	0.00	-0.03	0.16	0.00	-0.03
Borjalilu et al 2019	0.43	0.80	-0.25	0.46	0.84	-0.25	0.49	0.90	-0.25
Bosso, 2020	0.56	0.64	-0.23	0.58	0.66	-0.23	0.61	0.69	-0.23
Bostock et al 2019	0.33	-0.01	0.47	0.35	-0.01	0.47	0.37	-0.01	0.47
Carissoli et al 2015	-0.04	-0.06	0.13	-0.04	-0.06	0.13	-0.05	-0.07	0.13
Flett et al 2019a	0.14	-0.16	0.17	0.15	-0.17	0.17	0.16	-0.18	0.17
Flett et al 2019b	0.35	-0.16	0.12	0.36	-0.17	0.12	0.39	-0.18	0.12
Forbes 2020a	0.23	0.03	0.28	0.25	0.03	0.28	0.26	0.04	0.28
Forbes 2020b	0.23	-0.18	0.25	0.25	-0.19	0.25	0.26	-0.21	0.25
Howells et al 2016	0.29	0.07	0.36	0.31	0.07	0.36	0.33	0.08	0.36
Huberty et al 2022	0.45	0.19	0.41	0.48	0.20	0.41	0.51	0.22	0.41
Keng et al 2022	0.22	0.03	0.44	0.23	0.03	0.44	0.24	0.03	0.44
Kubo et al 2019	0.22	-0.02	0.50	0.23	-0.03	0.50	0.25	-0.03	0.50
Kubo et al 2019	0.23	0.00	0.09	0.24	0.00	0.09	0.26	0.00	0.09
Laird et al., 2022	-0.15	-0.24	-0.22	-0.16	-0.26	-0.22	-0.17	-0.28	-0.22
Lee & Jung, 2018	0.36	0.11	0.23	0.38	0.12	0.23	0.41	0.13	0.23
Leng et al. 2023	0.83	0.18	0.87	0.87	0.19	0.87	0.92	0.20	0.87
Levin et al., 2022	0.96	0.02	0.48	1.02	0.02	0.48	1.08	0.02	0.48
Luangapichart et al., 2022	1.13	0.58	0.44	1.20	0.61	0.44	1.28	0.65	0.44

Ly et al., 2014	0.89	1.35	-0.25	0.94	1.43	-0.25	1.01	1.52	-0.25
Nolan, 2020	0.17	0.26	0.23	0.18	0.28	0.23	0.19	0.29	0.23
Pratt et al. 2023	0.53	0.16	0.52	0.56	0.16	0.52	0.60	0.17	0.52
Rocamora González et al., 2022	-0.22	-0.41	-0.22	-0.23	-0.43	-0.22	-0.25	-0.46	-0.22
Smith et al 2021	0.30	0.02	0.44	0.32	0.02	0.44	0.34	0.03	0.44
Sun et al 2021	0.33	-0.10	0.27	0.34	-0.11	0.27	0.37	-0.11	0.27
Sun et al 2022	1.09	0.81	0.41	1.16	0.86	0.41	1.24	0.92	0.41
Treves 2023a	0.12	0.07	-0.04	0.12	0.07	-0.04	0.13	0.08	-0.04
Treves 2023b	0.13	0.08	-0.02	0.14	0.08	-0.02	0.15	0.09	-0.02
Treves 2023c	0.12	0.09	-0.14	0.12	0.09	-0.14	0.13	0.10	-0.14
Treves 2023d	0.13	0.08	0.03	0.14	0.08	0.03	0.15	0.09	0.03
Versluis et al 2018a	0.30	0.10	-0.12	0.30	0.11	-0.12	0.31	0.12	-0.12
Versluis et al 2018b	0.30	-0.04	0.08	0.30	-0.05	0.08	0.31	-0.05	0.08
Yoon et al 2022	0.37	-0.10	0.50	0.39	-0.11	0.50	0.42	-0.11	0.50

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores

Table S7 – depression as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .4 to .6

	r = .4			*r* = .5			*r* = .6		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Bear et al 2022	0.61	0.17	0.51	0.66	0.18	0.51	0.74	0.21	0.51
Bhayee et al 2016	0.18	0.00	-0.03	0.19	0.00	-0.03	0.21	0.00	-0.03
Borjalilu et al 2019	0.53	0.97	-0.25	0.57	1.06	-0.25	0.64	1.18	-0.25
Bosso, 2020	0.64	0.72	-0.23	0.67	0.76	-0.23	0.72	0.80	-0.23
Bostock et al 2019	0.40	-0.01	0.47	0.44	-0.01	0.47	0.49	-0.02	0.47
Carissoli et al 2015	-0.05	-0.07	0.13	-0.06	-0.08	0.13	-0.06	-0.09	0.13
Flett et al 2019a	0.17	-0.19	0.17	0.19	-0.21	0.17	0.21	-0.23	0.17
Flett et al 2019b	0.41	-0.19	0.12	0.45	-0.21	0.12	0.49	-0.23	0.12
Forbes 2020a	0.28	0.04	0.28	0.31	0.04	0.28	0.35	0.05	0.28

Forbes 2020b	0.28	-0.22	0.25	0.31	-0.24	0.25	0.35	-0.27	0.25
Howells et al 2016	0.35	0.08	0.36	0.38	0.09	0.36	0.42	0.10	0.36
Huberty et al 2022	0.55	0.24	0.41	0.60	0.26	0.41	0.67	0.29	0.41
Keng et al 2022	0.26	0.04	0.44	0.28	0.04	0.44	0.31	0.05	0.44
Kubo et al 2019	0.27	-0.03	0.50	0.29	-0.03	0.50	0.32	-0.04	0.50
Kubo et al 2019	0.28	0.00	0.09	0.31	0.00	0.09	0.35	0.00	0.09
Laird et al., 2022	-0.19	-0.30	-0.22	-0.20	-0.33	-0.22	-0.23	-0.36	-0.22
Lee & Jung, 2018	0.44	0.14	0.23	0.49	0.15	0.23	0.54	0.17	0.23
Leng et al., 2023	0.98	0.21	0.87	1.06	0.23	0.87	1.16	0.25	0.87
Levin et al., 2022	1.16	0.02	0.48	1.27	0.02	0.48	1.40	0.02	0.48
Luangapichart et al., 2022	1.38	0.70	0.44	1.52	0.77	0.44	1.69	0.85	0.44
Ly et al., 2014	1.09	1.63	-0.25	1.19	1.78	-0.25	1.33	1.97	-0.25
Nolan, 2020	0.21	0.32	0.23	0.23	0.35	0.23	0.26	0.39	0.23
Pratt et al., 2023	0.65	0.19	0.52	0.71	0.20	0.52	0.79	0.22	0.52
Rocamora González et al., 2022	-0.27	-0.49	-0.22	-0.29	-0.53	-0.22	-0.33	-0.59	-0.22
Smith et al 2021	0.37	0.03	0.44	0.40	0.03	0.44	0.45	0.03	0.44
Sun et al 2021	0.40	-0.12	0.27	0.44	-0.13	0.27	0.49	-0.14	0.27
Sun et al 2022	1.34	0.99	0.41	1.46	1.08	0.41	1.63	1.21	0.41
Treves 2023a	0.14	0.08	-0.04	0.16	0.09	-0.04	0.18	0.10	-0.04
Treves 2023b	0.16	0.10	-0.02	0.17	0.10	-0.02	0.19	0.12	-0.02
Treves 2023c	0.14	0.11	-0.14	0.16	0.12	-0.14	0.18	0.13	-0.14
Treves 2023d	0.16	0.09	0.03	0.17	0.10	0.03	0.19	0.11	0.03
Versluis et al 2018a	0.32	0.13	-0.12	0.34	0.14	-0.12	0.35	0.15	-0.12
Versluis et al 2018b	0.32	-0.05	0.08	0.34	-0.06	0.08	0.35	-0.06	0.08
Yoon et al 2022	0.45	-0.12	0.50	0.49	-0.13	0.50	0.55	-0.15	0.50

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores

Table S8– depression as outcome (+ve SMD = symptom improvement), repeated measures correlation estimates of .7 to .9

	r = .7			*r* = .8			*r* = .9		
Study	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$	$SMD_{withinINT}$	$SMD_{withinCONT}$	$SMD_{between}$
Bear et al 2022	0.85	0.24	0.51	1.04	0.29	0.51	1.44	0.41	0.51
Bhayee et al 2016	0.24	0.00	-0.03	0.28	0.00	-0.03	0.36	0.00	-0.03
Borjalilu et al 2019	0.73	1.35	-0.25	0.89	1.62	-0.25	1.21	2.19	-0.25
Bosso, 2020	0.77	0.85	-0.23	0.83	0.91	-0.23	0.92	0.98	-0.23
Bostock et al 2019	0.57	-0.02	0.47	0.69	-0.02	0.47	0.97	-0.03	0.47
Carissoli et al 2015	-0.07	-0.10	0.13	-0.08	-0.12	0.13	-0.11	-0.17	0.13
Flett et al 2019a	0.25	-0.25	0.17	0.30	-0.29	0.17	0.43	-0.35	0.17
Flett et al 2019b	0.55	-0.25	0.12	0.64	-0.29	0.12	0.79	-0.35	0.12
Forbes 2020a	0.40	0.06	0.28	0.49	0.07	0.28	0.69	0.09	0.28
Forbes 2020b	0.40	-0.31	0.25	0.49	-0.37	0.25	0.69	-0.50	0.25
Howells et al 2016	0.48	0.11	0.36	0.58	0.14	0.36	0.78	0.19	0.36
Huberty et al 2022	0.77	0.33	0.41	0.93	0.41	0.41	1.28	0.58	0.41
Keng et al 2022	0.35	0.05	0.44	0.41	0.06	0.44	0.51	0.09	0.44
Kubo et al 2019	0.37	-0.04	0.50	0.44	-0.05	0.50	0.58	-0.07	0.50
Kubo et al 2019	0.40	0.00	0.09	0.49	0.00	0.09	0.69	0.00	0.09
Laird et al., 2022	-0.26	-0.42	-0.22	-0.32	-0.51	-0.22	-0.45	-0.70	-0.22
Lee & Jung, 2018	0.63	0.20	0.23	0.76	0.24	0.23	1.07	0.34	0.23
Leng et al., 2023	1.30	0.29	0.87	1.49	0.33	0.87	1.82	0.42	0.87
Levin et al., 2022	1.59	0.03	0.48	1.89	0.03	0.48	2.47	0.05	0.48
Luangapichart et al., 2022	1.96	0.98	0.44	2.39	1.18	0.44	3.38	1.59	0.44
Ly et al., 2014	1.54	2.24	-0.25	1.88	2.67	-0.25	2.64	3.49	-0.25
Nolan, 2020	0.29	0.45	0.23	0.36	0.55	0.23	0.50	0.77	0.23
Pratt et al., 2023	0.91	0.25	0.52	1.11	0.28	0.52	1.54	0.35	0.52
Rocamora González et al., 2022	-0.38	-0.66	-0.22	-0.46	-0.77	-0.22	-0.65	-0.96	-0.22
Smith et al 2021	0.51	0.04	0.44	0.62	0.05	0.44	0.84	0.07	0.44
Sun et al 2021	0.56	-0.16	0.27	0.68	-0.18	0.27	0.96	-0.23	0.27
Sun et al 2022	1.87	1.38	0.41	2.27	1.68	0.41	3.12	2.29	0.41
Treves 2023a	0.20	0.11	-0.04	0.24	0.14	-0.04	0.34	0.19	-0.04
Treves 2023b	0.22	0.14	-0.02	0.27	0.17	-0.02	0.36	0.23	-0.02
Treves 2023c	0.20	0.15	-0.14	0.24	0.19	-0.14	0.34	0.26	-0.14

Treves 2023d	0.22	0.13	0.03	0.27	0.16	0.03	0.36	0.22	0.03
Versluis et al 2018a	0.36	0.18	-0.12	0.38	0.21	-0.12	0.40	0.29	-0.12
Versluis et al 2018b	0.36	-0.07	0.08	0.38	-0.09	0.08	0.40	-0.12	0.08
Yoon et al 2022	0.62	-0.17	0.50	0.75	-0.21	0.50	0.99	-0.28	0.50

Notes. Shading indicates between group effect is larger (more positive) than within group effect, *r* = estimated correlation between baseline and post scores