Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Beyond the Average Effect: A Meta-Analysis of Digital Cognitive Behavioural Therapy RCTS for Depression and Anxiety

Moses D. Tingir

mosestingir33@gmail.com

University of Mkar https://orcid.org/0009-0003-8747-673X

Blessing Iveren Yimam

University of Benin https://orcid.org/0009-0000-1291-2312

Maria Theresa Saliu

Northumbria Healthcare NHS Foundation Trust

Research Article

Keywords: Digital cognitive behavioural therapy, Internet-delivered CBT, Depression, Anxiety, Randomized controlled trials, Meta-analysis, Treatment heterogeneity, Digital mental health

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10842027/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Digital cognitive behavioural therapy (dCBT) has moved from a supplementary self-help option to a clinically relevant route for widening access to evidence-based psychological treatment. Its effects, however, vary across delivery formats, diagnostic groups, levels of human support and comparator conditions. This meta-analytic review estimated the post-treatment effects of dCBT on depressive and anxiety symptoms in randomized controlled trials published and interpreted the pooled estimates alongside recent implementation evidence. A focused, reproducible evidence search identified randomized trials in which a digital intervention had a clear CBT core, a randomized comparison condition, and an extractable standardized depression and/or anxiety contrast. Eight contrasts were synthesized for depression and eight for anxiety. Reported Cohen d values were converted to Hedges g; where necessary, standardized effects were derived from published group means and standard deviations. Random-effects models were estimated with restricted maximum likelihood and Hartung–Knapp confidence intervals. For depression, dCBT produced a moderate pooled benefit ($g = 0.49$, 95% CI 0.32–0.66; $I^2 = 54.6\%$; $k = 8$). For anxiety, the pooled effect was larger ($g = 0.70$, 95% CI 0.39–1.01; $I^2 = 84.6\%$; $k = 8$), but the prediction interval crossed zero. Removing two high-effect, GAD-specialist programmes reduced the anxiety estimate to $g = 0.53$ and I^2 to approximately 0%. Excluding the high-attrition Colombian depression trial similarly reduced depression heterogeneity while retaining a clinically meaningful effect. Contemporary dCBT generally improves depression and anxiety, but the magnitude of benefit depends materially on programme specificity, adherence, guidance, comparator stringency and population. Future trials should prioritize active comparators, longer follow-up, standardized engagement reporting and transparent safety monitoring.

1.0 Introduction

Depression and anxiety are among the conditions for which the gap between treatment need and treatment availability is especially visible. Evidence-based psychotherapy is effective, yet access is constrained by cost, travel, waiting lists, geographical maldistribution of clinicians, stigma, scheduling problems and the simple fact that conventional care depends heavily on professional time. Digital delivery does not remove all of these constraints, but it changes the unit of delivery. A therapeutic lesson, activity-monitoring task or exposure exercise can be made available repeatedly and asynchronously, while clinician input can be concentrated where it adds the most value. This scalability argument has driven rapid growth in internet- and mobile-based psychological interventions and, in particular, in digital forms of cognitive behavioural therapy (CBT) (Moshe et al., 2021; Linardon et al., 2024).

Digital CBT is not a single technology. Contemporary programmes range from structured web courses with weekly therapist messaging to app suites, brief self-guided behavioural activation tools and prescribed smartphone therapeutics. They also differ in whether they target a diagnosis, a transdiagnostic symptom cluster or a broader population such as university students. This variation matters because an effect attributed to “digital CBT” may actually reflect several interacting ingredients: the CBT content itself, the amount of human support, the credibility of the comparator, the intensity of prompting, the opportunity to practise skills in daily life, and participants’ willingness to persist with the programme (Bernstein et al., 2022; Boucher & Raiker, 2024).

Recent randomized trials illustrate both the promise and the complexity. In primary care, a coached mobile platform produced faster improvements in depression and anxiety than a treatment-as-usual wait-list condition (Graham et al., 2020). A pragmatic trial embedded in psychological services also found clinically important advantages for digital treatment, including better recovery and reliable improvement at eight weeks (Richards et al., 2020). At the same time, a university-student app trial showed a modest depression benefit and a less certain anxiety effect at the primary endpoint (McCloud et al., 2020), while a culturally adapted Colombian programme reported a large depression effect but suffered very high post-treatment missingness (Salamanca-Sanabria et al., 2020). These studies caution against reducing the literature to a single average without examining how that average is formed.

The evidence base has also changed substantially since 2020. Trials now include active psychoeducation comparators, blinded evaluator outcomes, integrated primary-care delivery, factorial testing of discrete CBT skills, culturally tailored programmes and data-driven platforms. The DaylightRx trial, for example, compared a smartphone-delivered dCBT programme with digital psychoeducation rather than no treatment and still reported a large anxiety effect at ten weeks (Parsons et al., 2025). The RESiLIENT programme used a large factorial design to test combinations of smartphone CBT skills for subthreshold depression in community adults, demonstrating the feasibility of decomposing digital psychotherapy into its active components rather than treating the intervention as a black box (Furukawa et al., 2025).

Meta-analytic evidence is broadly supportive but not uniform. Moshe et al. (2021) found meaningful effects across digital depression interventions, while Karyotaki et al. (2021) showed that guidance can be especially relevant for people with more severe depressive symptoms. Linardon et al. (2024), synthesizing 176 randomized trials of mental-health smartphone apps, confirmed benefits for depression and anxiety but also documented extensive heterogeneity. Bae et al. (2023) likewise found app-based interventions helpful in moderate-to-severe depression, and Terhorst et al. (2024) highlighted that average effects do not necessarily imply large patient-by-treatment heterogeneity. Taken together, the literature supports efficacy but leaves important questions about the size, stability and transportability of effects in newer RCTs. Accordingly, this article provides an updated, focused meta-analysis of randomized controlled trials published from 2020 onward.

1.1 Objectives of the study

This study has three objectives:

I. to estimate separate pooled effects for depressive and anxiety symptoms;
II. to examine heterogeneity through transparent sensitivity analyses;
III. to integrate the numerical findings with contemporary evidence on mechanisms, guidance, engagement, cultural adaptation and real-world implementation.

The emphasis is not on presenting dCBT as a replacement for all face-to-face care. Rather, it is on identifying what the current trial evidence justifies, where uncertainty remains, and which design choices appear most consequential for future clinical and research use.

2.0 Literature Review

2.1 Conceptual Review

2.1.1 Meaning and Scope of Digital Cognitive Behavioural Therapy

Digital cognitive behavioural therapy refers to CBT-informed treatment delivered primarily through a digital interface, such as a website, smartphone application or integrated digital platform. The defining feature is therefore therapeutic content rather than hardware. A programme can reasonably be described as dCBT when cognitive and behavioural change procedures form its central logic: identifying unhelpful appraisals, monitoring mood and behaviour, behavioural activation, graded exposure, problem solving, cognitive restructuring, worry management, relapse planning or structured practice. The intervention may be self-guided or supported by a therapist, coach or navigator. This distinction separates dCBT from generic wellness apps that provide relaxation or inspirational content without a coherent CBT treatment pathway (Jackson et al., 2023; Bernstein et al., 2022).

Delivery architecture affects how CBT is experienced. Web programmes often resemble sequential treatment modules, whereas mobile tools can distribute smaller exercises across everyday contexts. Graham et al. (2020) used an app-platform model in which participants could engage with a suite of brief tools; Bisby et al. (2024) tested a therapist-guided ultra-brief online treatment; and Dahne et al. (2025) evaluated a self-guided behavioural activation programme integrated into primary care. These models differ in intensity but preserve a common objective: convert evidence-based psychological procedures into repeatable, accessible actions that users can perform outside a consulting room.

A useful conceptual distinction is between digital delivery and digital therapeutic specificity. Some interventions digitize a conventional treatment pathway; others are designed around the affordances of mobile devices, including reminders, just-in-time practice, branching content or personalized sequencing. More recently, data-driven platforms have combined CBT with wearable data and navigation support (Fatouros et al., 2025). Such hybrid interventions may improve personalization, but they also make causal attribution harder because symptom change cannot be assigned to CBT alone without component-level testing.

2.1.2 Depression and Anxiety as Related but Distinct Targets

Depression and anxiety frequently co-occur, share cognitive and behavioural maintenance processes, and are commonly assessed together in digital-treatment research. Both may involve avoidance, reduced behavioural flexibility, perseverative thinking and negatively biased interpretations, yet the therapeutic emphasis is not identical. Depression-focused programmes often prioritize

behavioural activation, re-engagement with reinforcing activities, problem solving and modification of globally negative beliefs. Anxiety-focused programmes place greater weight on exposure, response prevention, worry management, intolerance of uncertainty and reappraisal of threat. A transdiagnostic platform can therefore address common processes while allowing symptom-specific modules to vary in depth.

The distinction matters analytically. A programme designed specifically for generalized anxiety disorder (GAD) may produce a larger anxiety effect than a broad programme offered to participants with mixed symptoms. Carl et al. (2020) reported a very large effect for a GAD-focused dCBT programme compared with wait-list, and Parsons et al. (2025) subsequently demonstrated a large benefit against a more demanding psychoeducation control. By contrast, broader university or primary-care programmes have typically yielded small-to-moderate or moderate standardized effects (McCloud et al., 2020; Graham et al., 2020). Pooling these trials without sensitivity analysis can obscure this clinically meaningful source of heterogeneity.

For depression, intervention specificity also varies. Salamanca-Sanabria et al. (2020) evaluated a culturally adapted depression programme for Colombian students, whereas Dahne et al. (2025) tested behavioural activation in primary care and Eylem-van Bergeijk et al. (2025) measured depression as a secondary outcome of a GAD app. The same label—depression outcome—therefore encompasses primary therapeutic targets and secondary symptom spillover. This review retains those outcomes when they are randomized and extractable but interprets them in light of treatment target and trial design.

2.1.3 Guidance, Engagement and Adherence

Human support can perform several functions in dCBT: clarify the rationale, reinforce practice, troubleshoot technology, encourage persistence, assess risk and provide accountability. It does not necessarily need to reproduce full psychotherapy. Bernstein et al. (2022) showed that app-based CBT studies have used a wide variety of support models, which makes “guided” versus “unguided” a coarse category. Karyotaki et al. (2021) further suggested that guidance may be particularly useful when depressive symptom severity is greater, whereas unguided approaches can still be effective for milder presentations.

Engagement is equally important but should not be treated as a simple dose count. Boucher and Raiker (2024) emphasized that engagement includes behavioural use and experiential involvement, and that retention problems are widespread in digital mental-health research. Jackson et al. (2023) argued that skill enactment and knowledge acquisition may be more theoretically informative than raw log-ins because CBT works through learning and application. The distinction explains why high session counts do not automatically imply clinical improvement, and why short interventions can work if they prompt focused practice of high-value skills.

Recent RCTs demonstrate these points. Graham et al. (2020) achieved high app use in a coached platform, but simple usage metrics were not strongly correlated with symptom change. In the Cerina feasibility trial, only a minority completed the recommended session dose, and intention-to-treat effects were smaller than completer effects (Eylem-van Bergeijk et al., 2025). Parsons et al. (2025), by contrast, observed a dose-response pattern between treatment engagement and GAD outcomes. These apparently different findings imply that engagement is not a universal moderator; its meaning depends on what is being counted and whether the counted activity reflects actual therapeutic practice.

Table 2.1
Core Concepts and Their Operational Meaning

Construct	Working definition	Relevance to this study	Sources
Digital CBT	CBT procedures delivered mainly via web/app/smartphone interfaces	Intervention must have a clear CBT core rather than generic wellness content	Jackson et al. (2023); Bernstein et al. (2022)
Guidance	Human support added to digital self-directed content	May improve accountability, tailoring and safety, especially with greater severity	Karyotaki et al. (2021); Bernstein et al. (2022)
Engagement	Behavioural use and meaningful participation in therapeutic tasks	Affects exposure to active ingredients but is not equivalent to log-in frequency	Boucher & Raiker (2024); Parsons et al. (2025)
Clinical effect	Between-group difference in symptom outcome at post-treatment	Standardized as Hedges g to allow scales to be combined	Harrer et al. (2021)
Implementation fit	Compatibility with setting, culture, workflow and user needs	Influences uptake, retention and transportability beyond efficacy trials	Salamanca-Sanabria et al. (2020); Dahne et al. (2025)

2.2 Theoretical Framework

The theoretical framework integrates cognitive, behavioural and self-regulation pathways. At the cognitive level, dCBT is expected to help users notice automatic interpretations, evaluate evidence, generate more balanced alternatives and reduce repetitive negative thinking. At the behavioural level, change occurs when users approach rather than avoid difficult situations, schedule rewarding activity, test predictions and practise coping responses. Digital delivery adds a self-regulation layer: reminders, monitoring, feedback and repeated access can place CBT practice closer to the contexts in which symptoms occur. The treatment effect therefore depends not only on information exposure but on whether digital interaction produces skill enactment.

This framework is consistent with process-oriented reviews. Domhardt et al. (2021) found that therapeutic processes such as working alliance, adherence and treatment expectancy are relevant within digital anxiety interventions, although the evidence is not equally developed for every mediator. Jackson et al. (2023) showed that digital CBT can improve both knowledge acquisition and use of therapeutic skills, supporting a learning-based mechanism rather than a passive information model. The framework also accommodates guidance: a coach may strengthen expectancy, task completion and troubleshooting without being the primary source of therapeutic content.

The model further predicts heterogeneity. A highly specific GAD programme that repeatedly trains relevant skills should, in principle, generate a stronger anxiety signal than a broad low-intensity programme used by a heterogeneous sample. Conversely, a culturally mismatched or poorly retained programme may fail before the intended cognitive or behavioural mechanisms are sufficiently activated. Hence, variability in pooled effects is theoretically meaningful rather than merely statistical noise.

2.3 Empirical Review

Table 2.2
Empirical Review of Recent Digital CBT and Closely Related Randomized Trials

Study	Population/sample	Intervention	Comparator	Key outcome	Interpretive note
Salamanca-Sanabria et al. (2020)	Colombian university students; N = 214	Culturally adapted 7-module iCBT + support	Wait-list	Depression d = 0.91; anxiety d = 0.60	Large effects, but very high missingness limits certainty.
Graham et al. (2020)	US primary care; N = 146	Coached IntelliCare app platform	TAU/wait-list	Adjusted d = 0.64 depression; 0.55 anxiety	Moderate benefits; recovery odds also favoured intervention.
McCloud et al. (2020)	UK university students; N = 168	Self-guided CBT-based mobile app	Wait-list	d = 0.39 depression; 0.41 anxiety at 6 weeks	Depression significant; anxiety estimate less precise at primary endpoint.
Richards et al. (2020)	Routine psychological services; N = 361	Digital depression/anxiety treatment with support	Wait-list	d = 0.55 depression; 0.63 anxiety	Pragmatic effectiveness and recovery benefits; sustained follow-up.
Carl et al. (2020)	Adults with moderate-to-severe GAD; N = 256	Smartphone dCBT for GAD	Wait-list	Anxiety d = 1.43	Large disorder-specific anxiety effect.
Al-Alawi et al. (2021)	Adults with COVID-19-related symptoms	Therapist-guided online therapy	Self-help internet therapy	Both arms improved	Active-comparator evidence; supports role of guidance but unsuitable for simple inactive-control pooling.
Klos et al. (2021)	University students; N = 181	AI chatbot with CBT-informed interaction	Psychoeducation book	No clear between-group superiority	Important null/near-null comparator evidence against overgeneralization.
Fatori et al. (2023)	Pregnant women with depression	Smartphone-assisted brief online CBT	Randomized comparison	Depressive symptoms improved	Shows extension of digitally assisted CBT to maternal mental health.
Bisby et al. (2024)	Adults with depression/anxiety; N = 242	Therapist-guided ultra-brief online treatment	Standard treatment + wait-list	d = 0.41 depression; 0.53 anxiety vs wait-list	Ultra-brief treatment effective at 9 weeks and broadly non-inferior to longer treatment.
Horwitz et al. (2024)	Adults seeking psychiatric services	Three digital interventions	Active digital comparators	Similar symptom improvement across arms	Demonstrates importance of comparator stringency and common digital-care factors.
Stiles-Shields et al. (2024)	Primary care; pilot N = 130	Technology-enabled service with sensing/support	Digital psychoeducation	No differential PHQ-9/GAD-7 trajectory	Active control reduced apparent incremental symptom advantage.
Parsons et al. (2025)	Adults with interview-confirmed GAD; N = 351	Self-directed tailored smartphone DCBT	Digital psychoeducation	d = 1.09 anxiety; d = 0.48 depression at 10 weeks	Large anxiety effect persisted against active control and to 24 weeks.
Dahne et al. (2025)	US primary care; N = 649	Moodivate behavioural activation app ± EHR integration	Usual care	Faster BDI-II improvement; derived endpoint d ≈ 0.39	Large pragmatic sample; digital arms increased odds of improvement/remission.
Eylem-van Bergeijk et al. (2025)	University students; N = 158	Unguided Cerina CBT app for GAD	Wait-list	ITT d ≈ 0.36 anxiety; d ≈ 0.17 depression	Promising feasibility; low session completion and smaller ITT than completer effects.
Furukawa et al. (2025)	Community adults with subthreshold	Smartphone CBT skill combinations	Factorial controls	Small-to-moderate component effects	Very large factorial programme identifies

Study	Population/sample	Intervention	Comparator	Key outcome	Interpretive note
	depression; N = 3,936				skill-level effects and excellent follow-up.
Fatouros et al. (2025)	Adults with depressive/GAD symptoms; N = 200	Data-driven CBT-centred digital therapeutic + wearable/navigator	Wait-list	Large reported symptom associations/effects	Innovative multimodal design; causal attribution to CBT alone is difficult.

2.4 Gap in Literature

Despite rapid expansion of digital mental-health trials, four gaps remain. First, recent meta-analyses combine broad app categories, whereas clinically meaningful differences can arise between CBT-specific programmes, generic digital interventions and multimodal therapeutics (Linardon et al., 2024; Bae et al., 2023). Second, comparator conditions vary from wait-list to credible psychoeducation, making pooled efficacy sensitive to control stringency (Horwitz et al., 2024; Parsons et al., 2025). Third, engagement is inconsistently operationalized, so the relationship between programme use, skill enactment and outcome remains difficult to compare across trials (Jackson et al., 2023; Boucher & Raiker, 2024). Fourth, highly positive effects coexist with substantial attrition in some studies, including the culturally adapted Colombian trial (Salamanca-Sanabria et al., 2020). An updated synthesis therefore needs separate depression and anxiety estimates, transparent treatment of extractable RCT data, and sensitivity analyses that show whether high-effect specialist programmes or high-risk attrition patterns are driving the average. This review addresses that gap using post-2020 randomized evidence.

3.0 Methodology

3.1 Review Design and Reporting Approach

This study was designed as a focused meta-analytic review of randomized controlled trials published from 2020. Reporting was informed by PRISMA 2020 principles, particularly explicit eligibility criteria, transparent study flow and separation of qualitative from quantitative evidence (Page et al., 2021). It is important to distinguish the present work from a protocol-registered, exhaustive systematic review: the evidence search was structured and reproducible but used PubMed-indexed web results, publisher repositories and backward/forward citation chaining rather than a licensed multi-database export with deduplicated record counts. Accordingly, no invented database totals or screening counts are presented.

Primary trial reports were preferred when extracting sample sizes, outcome definitions and standardized effects. Recent systematic reviews and meta-analyses were used to identify potentially relevant RCTs, triangulate interpretation and locate implementation themes, but they did not contribute duplicate effect sizes to the quantitative model (Moshe et al., 2021; Linardon et al., 2024; Terhorst et al., 2024). The unit of analysis was one eligible intervention-versus-comparator contrast per study and outcome at the primary or most comparable post-treatment endpoint.

3.2 Eligibility Criteria

Table 3.1
PICOS-Based Eligibility Criteria

Domain	Criterion
Population	Adults or university-age participants with depressive and/or anxiety symptoms, including diagnosed GAD and primary-care populations.
Intervention	Digital intervention delivered primarily through web/app/smartphone technology with a clear CBT, behavioural activation, exposure, cognitive restructuring or related CBT skill core.
Comparator	Randomized wait-list, treatment as usual/usual care, or credible active digital psychoeducation. Multi-arm trials contributed only one independent comparison to avoid double counting a shared control.
Outcomes	Validated depressive and/or anxiety symptom measure with a standardized between-group effect or sufficient post-treatment data to derive one.
Design	Randomized controlled trial; 2020 onward; peer-reviewed primary report. Pilot RCTs could be included when randomized and extractable.
Exclusion	Single-arm studies; observational cohorts; protocols; non-CBT wellness programmes; duplicate/secondary analyses without an independent extractable contrast; factorial or multimodal designs where a single CBT-versus-control effect could not be isolated.

3.3 Search and Study Identification

Search concepts combined terms for digital or internet-delivered CBT with depression, anxiety and randomized trials. Practical combinations included phrases such as “digital cognitive behavioral therapy depression randomized trial,” “internet CBT anxiety randomized controlled trial,” “mobile CBT depression anxiety RCT,” and title/DOI verification searches. PubMed records and primary publisher pages were prioritized for factual extraction. Recent meta-analyses and systematic reviews were also citation-chained to identify trials that might use alternative intervention labels such as behavioural activation, coached app platform or technology-enabled service.

Eligibility was applied in two stages. First, a broader empirical-review set was assembled to map recent trial designs and comparator types. Second, the quantitative set was restricted to trials with an independent randomized contrast and an extractable standardized post-treatment effect. This prevented several attractive but analytically incompatible studies from being forced into the pool. For example, Furukawa et al. (2025) used four factorial trials to estimate effects of combinations of CBT skills; Fatouros et al. (2025) combined CBT with wearables and navigation; and Horwitz et al. (2024) compared active digital interventions without a clear inactive reference. These studies remained important for interpretation but were not used to manufacture a single conventional effect size.

3.4 Data Extraction and Effect-Size Harmonization

For each quantitatively eligible trial, the extraction fields were author/year, randomized arm sizes, intervention type, comparator, outcome, endpoint, reported standardized effect and any confidence interval. Positive values were coded to mean lower symptoms in the dCBT group relative to the comparator. Cohen d estimates were converted to Hedges g using the small-sample correction. When a published 95% confidence interval for d was available, its standard error was reconstructed from the interval and then corrected consistently. When only group means and standard deviations were available, a post-treatment standardized mean difference was calculated.

One such derivation was required for the primary Moodivate-versus-usual-care contrast reported by Dahne et al. (2025). The published week-12 BDI-II values were 21.67 (SD 13.03) for Moodivate and 27.04 (SD 14.32) for usual care; these values yield a between-group Cohen d of approximately 0.39 before the small-sample correction. This derived contrast is deliberately more conservative than interpreting the trial’s within-arm change effect sizes as if they were between-group effects. For Eylem-van Bergeijk et al. (2025), the published intention-to-treat standardized effects from the trial table were used rather than the larger completer-only estimates.

In multi-arm studies, only one dCBT arm was paired with a shared control in a given pool. For Bisby et al. (2024), the ultra-brief treatment was compared with wait-list; the standard-length active treatment was retained for narrative interpretation. For Dahne et al. (2025), Moodivate without EHR integration was compared with usual care. This rule avoided artificially increasing precision by counting the same control participants twice.

3.5 Statistical Analysis

Separate random-effects models were estimated for depression and anxiety. Restricted maximum likelihood (REML) was used for between-study variance (τ^2), and Hartung–Knapp adjustment was used for the confidence interval around the pooled mean because the number of studies was small and heterogeneity could not be assumed negligible. The approach follows contemporary meta-analytic recommendations that emphasize uncertainty around τ^2 rather than relying only on a normal approximation (Harrer et al., 2021). Statistical heterogeneity was summarized with Cochran Q, τ^2 and I^2. A 95% prediction interval was also calculated to describe the approximate range in which the underlying effect of a comparable future study could fall.

Two sensitivity analyses were prespecified after inspection of design features rather than selected solely on statistical grounds. The depression model was re-estimated without Salamanca-Sanabria et al. (2020) because post-treatment missingness was exceptionally high. The anxiety model was re-estimated after removing the two specialist GAD programmes with very large effects—Carl et al. (2020) and Parsons et al. (2025)—to determine whether broad dCBT trials produced a more homogeneous underlying effect. Because each outcome contained fewer than ten studies, formal meta-regression and funnel-plot asymmetry tests were not treated as reliable and were not used to make claims about moderators or publication bias.

4.0 Results

4.1 Characteristics of the Quantitative Evidence

Eight independent randomized contrasts contributed to at least one quantitative model. The depression pool contained eight contrasts and an approximate pairwise randomized denominator of 1,988 participants after inclusion of the Cerina secondary depression outcome; the anxiety pool contained eight contrasts and 1,816 randomized participants. Trial settings included university populations, primary care, routine psychological services, general community recruitment and diagnosis-specific GAD samples. Comparators were predominantly wait-list or usual care, with the Parsons et al. (2025) GAD trial providing an important active psychoeducation contrast. Treatment duration ranged from ultra-brief formats to approximately 8–12 weeks for the primary endpoint.

Table 4.1
Effect-Size Inputs Used in the Quantitative Synthesis

Study	dCBT/control n	Comparator	Outcome	Standardized input	Extraction note
Salamanca-Sanabria et al. (2020)	107/107	Wait-list	Depression	0.91	Reported d; very high missingness
Salamanca-Sanabria et al. (2020)	107/107	Wait-list	Anxiety	0.60	Reported d
Graham et al. (2020)	74/72	TAU/wait-list	Depression	0.64	Adjusted week-8 d
Graham et al. (2020)	74/72	TAU/wait-list	Anxiety	0.55	Adjusted week-8 d
McCloud et al. (2020)	84/84	Wait-list	Depression	0.39	Adjusted week-6 standardized effect
McCloud et al. (2020)	84/84	Wait-list	Anxiety	0.41	Adjusted week-6 standardized effect
Richards et al. (2020)	241/120	Wait-list	Depression	0.55 [0.32, 0.77]	Reported d and CI
Richards et al. (2020)	241/120	Wait-list	Anxiety	0.63 [0.40, 0.85]	Reported d and CI
Carl et al. (2020)	128/128	Wait-list	Anxiety	1.43	GAD-specific programme
Bisby et al. (2024)	85/77	Wait-list	Depression	0.41	Ultra-brief vs wait-list
Bisby et al. (2024)	85/77	Wait-list	Anxiety	0.53	Ultra-brief vs wait-list
Parsons et al. (2025)	175/176	Psychoeducation	Depression	0.48 [0.29, 0.67]	Secondary PHQ-8 endpoint
Parsons et al. (2025)	175/176	Psychoeducation	Anxiety	1.09 [0.81, 1.37]	Primary GAD-7 endpoint
Dahne et al. (2025)	216/212	Usual care	Depression	≈ 0.39	Derived from week-12 mean/SD
Eylem-van Bergeijk et al. (2025)	79/79	Wait-list	Depression	0.17	Absolute ITT table d
Eylem-van Bergeijk et al. (2025)	79/79	Wait-list	Anxiety	0.36	Absolute ITT table d

4.2 Pooled Effect on Depressive Symptoms

Across eight randomized contrasts, dCBT was associated with lower depressive symptoms than comparator conditions (Hedges g = 0.49, 95% CI 0.32–0.66, Hartung–Knapp p < .001). Heterogeneity was moderate ($Q(7) = 15.43$, $p = 0.031$; $I^2 = 54.6\%$; $\tau^2 = 0.022$). The 95% prediction interval ranged from 0.09 to 0.89, remaining on the beneficial side of zero but indicating appreciable variation in the magnitude likely across settings. The point estimate corresponds to a moderate standardized advantage and is not dependent on a single very large contemporary primary-care trial.

The prespecified sensitivity analysis removed the Salamanca-Sanabria et al. (2020) trial because approximately four-fifths of the intervention arm lacked post-treatment data. The pooled effect remained statistically and clinically meaningful at g = 0.44 (95% CI 0.32–0.56), while I^2 fell to 0.0%. This pattern suggests that the principal conclusion—benefit for depressive symptoms—is robust, but the larger Colombian estimate contributed substantially to between-study dispersion. The result also illustrates why attrition should be treated as an interpretive feature rather than a footnote in digital trials.

4.3 Pooled Effect on Anxiety Symptoms

The anxiety model also favoured dCBT, with a larger pooled mean than the depression model (Hedges g = 0.70, 95% CI 0.39–1.01; p = 0.001). Statistical heterogeneity was high ($Q(7) = 45.39$, $p < .001$; $I^2 = 84.6\%$; $\tau^2 = 0.115$). The 95% prediction interval (-0.19 to 1.59) crossed zero, showing that a large average should not be interpreted as a guarantee of a large effect in every population or programme. The spread was visibly driven by the specialist GAD trials, particularly Carl et al. (2020) and Parsons et al. (2025), whose effects exceeded those of broader student and primary-care interventions.

When the two specialist GAD programmes were removed, the pooled anxiety effect was g = 0.53 (95% CI 0.38–0.68) and I^2 was approximately 0.0%. The remaining trials therefore converged closely around a moderate effect. This sensitivity analysis is not an argument that the GAD trials are outliers to be discarded; they represent clinically important evidence. Instead, it indicates that diagnosis-specific design, intervention intensity and possibly recruitment severity explain much of the apparent inconsistency in the broad anxiety pool.

Table 4.2
Random-Effects Meta-Analysis and Sensitivity Results

Model	k	Hedges g	95% CI	I^2	τ^2	95% prediction interval
Depression: primary model	8	0.49	0.32–0.66	54.6%	0.022	0.09–0.89
Depression: excluding high-attrition trial	7	0.44	0.32–0.56	0.0%	0.000	0.32–0.56
Anxiety: primary model	8	0.70	0.39–1.01	84.6%	0.115	-0.19–1.59
Anxiety: excluding 2 GAD-specialist trials	6	0.53	0.38–0.68	0.0%	0.000	0.36–0.69

4.4 Qualitative Risk-of-Bias Considerations

A formal domain-by-domain Cochrane RoB 2 adjudication was not imposed retrospectively without full protocol packages for every trial. Instead, this review used a transparent qualitative appraisal of features most likely to affect the pooled estimates: randomization, comparator credibility, outcome measurement, missingness and selective analytic flexibility. Randomization was generally well described in the large recent trials, but blinding of participants is rarely possible in behavioural digital interventions. Self-reported symptom scales therefore remain vulnerable to expectancy effects, particularly against wait-list. Active psychoeducation in Parsons et al. (2025) provides stronger protection against this problem. Missing data were the clearest concern in Salamanca-Sanabria et al. (2020), while Eylem-van Bergeijk et al. (2025) showed a marked difference between completer and intention-to-treat effects. Large pragmatic trials such as Dahne et al. (2025) reduce imprecision but still depend mainly on self-report symptom outcomes.

Table 4.3
Qualitative Risk-of-Bias and Evidence-Strength Appraisal

Study	Randomization	Comparator	Outcome mode	Missingness/adherence	Overall interpretive concern
Salamanca-Sanabria et al. (2020)	Randomized	Wait-list	Self-report	High missingness	High concern mainly from attrition
Graham et al. (2020)	Randomized	TAU/wait-list	Self-report	Managed with imputation/modeling	Some concern; inactive comparator
McCloud et al. (2020)	Randomized	Wait-list	Self-report	Notable attrition	Some concern
Richards et al. (2020)	Randomized	Wait-list	Self-report	Reasonable pragmatic retention	Some concern; inactive comparator
Carl et al. (2020)	Randomized	Wait-list	Self-report/clinical context	Acceptable	Some concern; large inactive-control effect
Bisby et al. (2024)	Randomized	Wait-list + active arm	Self-report	Reported follow-up	Some concern
Parsons et al. (2025)	Randomized; blinded evaluator	Active psychoeducation	Self-report + blinded clinician	High retention	Lower concern relative to most digital trials
Dahne et al. (2025)	Randomized	Usual care	Self-report	Large pragmatic sample	Some concern; treatment not blinded
Eylem-van Bergeijk et al. (2025)	Randomized pilot	Wait-list	Self-report	Low intervention completion	Some/high concern for adherence and pilot imprecision

5.0 Discussion

5.1 Principal Findings

The central finding is that contemporary dCBT RCTs provide consistent evidence of symptom benefit, but the size of that benefit is outcome- and context-dependent. The pooled depression effect was moderate ($g \approx 0.49$), and it remained meaningful after removing the trial with the most severe attrition problem. The anxiety average was larger ($g \approx 0.70$) but substantially more heterogeneous. Once the two diagnosis-specific GAD programmes were removed, the anxiety effect converged on approximately $g = 0.53$, closely resembling the magnitude seen in the broader depression pool. The most defensible general conclusion is therefore not that dCBT is "large-effect" treatment across the board, but that a moderate average benefit is robust and that more specialized programmes can sometimes achieve substantially larger gains.

This pattern agrees with prior meta-analytic work while sharpening its interpretation. Moshe et al. (2021) concluded that digital interventions can reduce depressive symptoms, and Karyotaki et al. (2021) showed that internet CBT is effective across severity levels while guidance may matter more for moderate-to-severe presentations. Linardon et al. (2024) similarly found positive effects across a very large set of smartphone trials but reported heterogeneity by intervention and study characteristics. The present review, restricted to newer RCTs and separate symptom domains, reproduces the broad efficacy signal without assuming that app modality or digital format is itself the active ingredient.

The anxiety results are especially informative. Carl et al. (2020) reported $d = 1.43$ for a GAD-focused programme, and Parsons et al. (2025) later demonstrated $d = 1.09$ against digital psychoeducation with benefits maintained to 24 weeks. Because the latter used an active comparator and blinded evaluator outcomes for remission, the large effect cannot be dismissed simply as a wait-list artifact. Nevertheless, the lower and highly homogeneous pooled effect among the remaining anxiety trials indicates that the Daylight family of

GAD interventions represents a more intensive, disorder-specific segment of the dCBT literature rather than a typical estimate for every anxiety app.

5.2 Comparator Conditions and the Meaning of Efficacy

Control-group design is one of the most consequential choices in digital psychotherapy research. Wait-list comparisons answer whether immediate access to dCBT is better than delayed access, but they do not control for treatment expectancy, structured attention, digital contact or self-monitoring. Active digital controls ask a stricter question: whether the CBT-specific programme adds benefit beyond credible information and engagement. The contrast between trial types should therefore be understood as different estimands rather than as stronger and weaker versions of the same experiment.

Parsons et al. (2025) is notable because its active psychoeducation control matched several nonspecific digital features yet dCBT still produced a large anxiety advantage. In contrast, Stiles-Shields et al. (2024) found no differential symptom trajectory when a technology-enabled service was compared with digital psychoeducation, and Horwitz et al. (2024) reported broadly similar outcomes across three active digital interventions for adults seeking psychiatric services. Klos et al. (2021) also found no clear superiority of an AI chatbot over psychoeducational material in a pilot university sample. These null or small incremental effects are not failures of the field; they reveal how much of improvement may be attributable to common digital-care factors and natural recovery when comparators are credible.

For future meta-analyses, comparator type should therefore be coded as a primary design variable, and trial reports should state whether the research question is efficacy against inactivity, incremental efficacy against a credible digital control, or effectiveness against usual care. Combining these questions can produce an average that is statistically valid but clinically ambiguous.

5.3 Guidance, Skill Enactment and Engagement

The evidence also complicates the simple claim that more guidance is always better. Guided programmes frequently perform well, and Karyotaki et al. (2021) provides strong evidence that guidance is useful for more severe depression. Bernstein et al. (2022) described human support as a potentially important component of app-based CBT, but support models vary widely in frequency, expertise and purpose. Bisby et al. (2024) further showed that a therapist-guided ultra-brief intervention can achieve moderate effects while using less treatment time than a conventional digital course. This supports a stepped-care view in which guidance is allocated according to need rather than treated as an all-or-none requirement.

Engagement findings are equally nuanced. Parsons et al. (2025) reported larger GAD effects among participants who completed more techniques and lessons, whereas Graham et al. (2020) found that simple app-use metrics were not strongly associated with symptom change. Eylem-van Bergeijk et al. (2025) documented a larger completer effect than intention-to-treat effect, suggesting that insufficient exposure limited population-level impact. These observations align with Jackson et al. (2023): what matters mechanistically is likely enactment of therapeutic skills, not merely opening the application.

This distinction has practical implications for digital product design. A programme that maximizes notifications and screen time may increase engagement statistics without increasing cognitive restructuring, exposure or behavioural activation. Conversely, a short exercise that leads a user to confront an avoided situation can represent high therapeutic engagement despite low digital dwell time. Future RCTs should therefore report both platform analytics and proximal indicators of CBT skill practice.

5.4 Cultural and Service-Delivery Context

Digital delivery can expand geographical reach, but it does not automatically solve cultural or service-fit problems. Salamanca-Sanabria et al. (2020) is important because it demonstrated a large effect after cultural adaptation for Colombian students, a setting underrepresented in earlier internet-CBT literature. Yet its severe attrition also shows that translation and adaptation alone do not guarantee sustained participation. Cultural fit, internet reliability, perceived privacy, treatment credibility and competing demands can all determine whether a theoretically sound programme is used long enough to matter.

Primary-care trials provide a different lesson. Graham et al. (2020) recruited a population that is often underserved in technology research, while Dahne et al. (2025) embedded a self-guided behavioural activation programme across 22 clinics and demonstrated improved depression trajectories and greater odds of clinically significant improvement and remission. The modest uptake of EHR functionality by clinicians in the Dahne trial is instructive: successful patient-facing technology does not guarantee clinician-facing workflow adoption. Implementation therefore has at least two levels patient engagement and organizational integration.

The wider literature supports this distinction. Boucher and Raiker (2024) identify retention as a recurring challenge across digital mental health, while Terhorst et al. (2024) caution against assuming that all observed variation reflects stable person-specific treatment effects. Service conditions, interface design and support may account for a substantial share of heterogeneity. For Scopus-indexed clinical research, implementation variables should be treated as explanatory constructs rather than appended as descriptive usage statistics.

5.5 Relation to Recent Meta-Analyses and Component Research

The magnitude of the present depression estimate is consistent with the direction of Bae et al. (2023), who synthesized app-based interventions for moderate-to-severe depression, and with the broader review by Moshe et al. (2021). Linardon et al. (2024) found that mental-health smartphone apps can reduce depression and anxiety across a much larger evidence base, but their broad app inclusion captures interventions that are not necessarily CBT-specific. Restricting the present analysis to a clear CBT core increases conceptual coherence while reducing the number of combinable trials. Longer-term evidence also indicates that transdiagnostic internet-delivered CBT can retain clinically meaningful benefits beyond the acute treatment window, although predictors of sustained response remain incompletely resolved (González-Robles et al., 2022).

Component research is likely the next major advance. Furukawa et al. (2025) used four 2×2 factorial trials involving nearly four thousand community adults to test smartphone CBT skills for subthreshold depression. That design moves beyond asking whether a branded programme works and toward identifying which skills or combinations produce benefit. Fatouros et al. (2025) represents another direction: multimodal personalization using sensor data and a digital navigator. These innovations are clinically attractive, but conventional two-arm meta-analysis cannot cleanly isolate the CBT contribution when several components change together.

Process evidence should therefore be linked to efficacy evidence. Domhardt et al. (2021) synthesized therapeutic processes in digital anxiety interventions, and Jackson et al. (2023) focused directly on skill enactment and knowledge acquisition. Their work suggests a future research model in which clinical outcome, engagement, skill use and treatment process are analyzed jointly. Such designs can explain why two programmes with similar CBT content produce different average effects and why high usage is sometimes weakly related to improvement.

6.0 Clinical, Policy and Design Implications

For clinical services, the evidence supports dCBT as a legitimate treatment pathway rather than merely an educational adjunct. A moderate average depression effect and a moderate-to-large anxiety effect are large enough to matter at population scale, particularly when conventional therapy capacity is limited. However, deployment should be matched to clinical presentation. Self-guided or lightly supported programmes may be reasonable first-line or stepped-care options for mild-to-moderate symptoms, while greater severity, comorbidity, risk or repeated nonresponse may justify more intensive professional involvement. Digital treatment should be integrated with escalation pathways rather than offered as an isolated application.

For health systems, implementation decisions should distinguish evidence of efficacy from evidence of workflow fit. The primary-care findings of Graham et al. (2020) and Dahne et al. (2025) show that digital interventions can operate in real clinical settings, but organizational adoption may lag behind patient availability. Procurement decisions should therefore examine clinical evidence, privacy, interoperability, safety monitoring, clinician burden, accessibility and engagement reporting. A platform with a strong RCT but poor integration can still fail at service level.

For designers, the results favour therapeutic specificity over feature accumulation. The very large GAD effects in Carl et al. (2020) and Parsons et al. (2025) arose from programmes engineered around disorder-relevant CBT skills, while broad programmes generally produced more moderate effects. This does not mean every app should be diagnosis-specific; transdiagnostic treatment remains useful. It means that the intended mechanism and target population should be explicit, and product analytics should measure whether users actually practise the targeted therapeutic behaviours.

7.0 Limitations and Directions for Future Research

This review has several limitations. Most importantly, it is a focused update meta-analysis rather than an exhaustive, protocol-registered systematic review exported from multiple bibliographic databases. The search strategy prioritized primary trial verification and transparent inclusion of extractable RCTs, but a licensed database search could identify additional small studies or non-English

reports. The article should therefore be interpreted as a reproducible synthesis of a contemporary RCT set, not as a definitive census of every digital CBT trial published since 2020.

Second, standardized effects were reported in different ways across trials. Some were adjusted model-based contrasts, some were raw standardized mean differences and one depression contrast was derived from published endpoint means and standard deviations. Hedges g places these on a common scale but cannot eliminate differences in estimands. Third, the number of studies per outcome was small, preventing reliable meta-regression. Apparent moderators such as guidance, comparator type or population severity are therefore discussed as plausible design explanations rather than statistically proven moderators.

Fourth, most outcomes were self-reported, and participant blinding is generally impossible in psychological treatment trials. Wait-list comparators may exaggerate treatment-specific effects because expectancy and structured engagement differ between groups. The active-control design and blinded evaluator outcomes used by Parsons et al. (2025) provide a stronger benchmark for future studies. Fifth, follow-up periods were inconsistent. Richards et al. (2020) and Parsons et al. (2025) provide evidence of sustained benefit, but many digital trials still focus on immediate post-treatment outcomes.

Future research should adopt four priorities. Trials should use credible active comparators when the aim is to isolate CBT-specific efficacy; report adverse events and deterioration in a standardized way; distinguish platform use from therapeutic skill enactment; and publish sufficient data to calculate between-group effects without reconstructing them from figures or within-arm change scores. Larger component and factorial designs, such as Furukawa et al. (2025), can identify which digital CBT ingredients are necessary. Studies in low- and middle-income settings should also move beyond cultural translation to test retention strategies, affordability, connectivity and implementation within local care pathways.

8.0 Conclusion

Randomized evidence published indicates that digital cognitive behavioural therapy can meaningfully reduce depressive and anxiety symptoms. In the present synthesis, the average depression benefit was moderate and remained robust after excluding a high-attrition study. Anxiety showed a larger pooled effect, but most of the heterogeneity was attributable to two high-effect, GAD-specialist programmes; the remaining trials converged on a moderate effect. These results support dCBT as an evidence-based route for expanding psychological treatment while cautioning against treating all digital interventions, populations and comparators as interchangeable.

The practical question is no longer simply whether therapy can be delivered through a screen. The more useful questions concern what CBT content is delivered, to whom, with what support, against what comparator, and whether users translate digital lessons into real-world therapeutic action. Future progress will depend on stronger active-control trials, component-level experimentation, transparent adherence and safety reporting, and implementation studies that connect efficacy with service realities.

References

1. Al-Alawi M, McCall RK, Sultan A, Al Balushi N, Al-Mahrouqi T, Al Ghailani A, Sabti A, Al-Maniri H, Panchatcharam A, S. M., Sinawi A, H (2021) Efficacy of a six-week-long therapist-guided online therapy versus self-help internet-based therapy for COVID-19-induced anxiety and depression: Open-label, pragmatic, randomized controlled trial. JMIR Mental Health 8(2):e26683. https://doi.org/10.2196/26683
2. Bae H, Shin H, Ji H-G, Kwon JS, Kim H, Hur J-W (2023) App-based interventions for moderate to severe depression: A systematic review and meta-analysis. JAMA Netw Open 6(11):e2344120. https://doi.org/10.1001/jamanetworkopen.2023.44120
3. Bernstein EE, Weingarden H, Wolfe EC, Hall MD, Snorrason I, Wilhelm S (2022) Human support in app-based cognitive behavioral therapies for emotional disorders: Scoping review. J Med Internet Res 24(4):e33307. https://doi.org/10.2196/33307
4. Bisby MA, Balakumar T, Scott AJ, Titov N, Dear BF (2024) An online therapist-guided ultra-brief treatment for depression and anxiety: A randomized controlled trial. Psychol Med 54(5):902–913. https://doi.org/10.1017/S003329172300260X
5. Boucher EM, Raiker JS (2024) Engagement and retention in digital mental health interventions: A narrative review. BMC Digit Health 2:52. https://doi.org/10.1186/s44247-024-00105-9
6. Carl JR, Miller CB, Henry AL, Davis ML, Stott R, Smits JAJ, Emsley R, Gu J, Shin O, Otto MW, Craske MG, Saunders KEA, Goodwin GM, Espie CA (2020) Efficacy of digital cognitive behavioral therapy for moderate-to-severe symptoms of generalized anxiety

disorder: A randomized controlled trial. Depress Anxiety 37(12):1168–1178. https://doi.org/10.1002/da.23079

7. Dahne J, Wahlquist AE, Carpenter MJ, Graboyes EM, Lejuez CW, Kustanowitz J, Natale N, Levins O, Player M, Diaz VA (2025) A digital depression treatment program for adults treated in primary care: A randomized clinical trial. JAMA Intern Med 185(6):692–701. https://doi.org/10.1001/jamainternmed.2025.0494
8. Domhardt M, Nowak H, Engler S, Baumel A, Grund S, Mayer A, Terhorst Y, Baumeister H (2021) Therapeutic processes in digital interventions for anxiety: A systematic review and meta-analytic structural equation modeling of randomized controlled trials. Clin Psychol Rev 90:102084. https://doi.org/10.1016/j.cpr.2021.102084
9. Eylem-van Bergeijk O, Robinson T, Manktelow M, Olympios M, Poulter S, Mane P, Panagioti M, Condell J, Leavey G (2025) Cerina—Cognitive behavioral therapy–based mobile app for managing generalized anxiety disorder symptoms among university students: Results from a pilot feasibility randomized controlled trial. JMIR mHealth uHealth 13:e70691. https://doi.org/10.2196/70691
10. Fatori D, Zuccolo P, Xavier MO, Matijasevich A, Polanczyk GV (2023) Smartphone-assisted online brief cognitive behavioral therapy to treat maternal depression: Findings of a randomized controlled trial. Brazilian J Psychiatry 45(1):50–53. https://doi.org/10.47626/1516-4446-2022-2679
11. Fatouros P, Tsirmpas C, Andrikopoulos D, Kaplow S, Kontoangelos K, Papageorgiou C (2025) npj Digit Med 8:113. https://doi.org/10.1038/s41746-025-01511-7. Randomized controlled study of a digital data driven intervention for depressive and generalized anxiety symptoms
12. Furukawa TA, Tajika A, Toyomoto R, Sakata M, Luo Y, Horikoshi M, Akechi T, Kawakami N, Nakayama T, Kondo N, Fukuma S, Kessler RC, Christensen H, Whitton A, Nahum-Shani I, Lutz W, Cuijpers P, Wason JMS, Noma H (2025) Cognitive behavioral therapy skills via a smartphone app for subthreshold depression among adults in the community: The RESiLIENT randomized controlled trial. Nat Med 31(6):1830–1839. https://doi.org/10.1038/s41591-025-03639-1
13. González-Robles A, Roca P, Díaz-García A, García-Palacios A, Botella C (2022) Long-term effectiveness and predictors of transdiagnostic internet-delivered cognitive behavioral therapy for emotional disorders in specialized care: Secondary analysis of a randomized controlled trial. JMIR Mental Health 9(10):e40268. https://doi.org/10.2196/40268
14. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, Mohr DC (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
15. Harrer M, Cuijpers P, Furukawa TA, Ebert DD (2021) Doing meta-analysis with R: A hands-on guide. Chapman & Hall/CRC. https://doi.org/10.1201/9781003107347
16. Horwitz AG, Mills ED, Sen S, Bohnert ASB (2024) Comparative effectiveness of three digital interventions for adults seeking psychiatric services: A randomized clinical trial. JAMA Netw Open 7(7):e2422115. https://doi.org/10.1001/jamanetworkopen.2024.22115
17. Jackson HM, Calear AL, Batterham PJ, Ohan JL, Farmer GM, Farrer LM (2023) Skill enactment and knowledge acquisition in digital cognitive behavioral therapy for depression and anxiety: Systematic review of randomized controlled trials. J Med Internet Res 25:e44673. https://doi.org/10.2196/44673
18. Karyotaki E, Efthimiou O, Miguel C, Maas Genannt Bermpohl F, Furukawa TA, Cuijpers P, Riper H, Patel V, Mira A, Gemmil AW, Yeung AS, Lange A, Williams AD, Mackinnon A, Geraedts A, Van Straten A, Meyer B, Björkelund C, Knaevelsrud C, Forsell Y (2021) Internet-based cognitive behavioral therapy for depression: A systematic review and individual patient data network meta-analysis. JAMA Psychiatry 78(4):361–371. https://doi.org/10.1001/jamapsychiatry.2020.4364
19. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL (2021) Artificial intelligence–based chatbot for anxiety and depression in university students: Pilot randomized controlled trial. JMIR Formative Res 5(8):e20678. https://doi.org/10.2196/20678
20. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M (2024) Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety: A meta-analysis of 176 randomized controlled trials. World Psychiatry 23(1):139–149. https://doi.org/10.1002/wps.21183
21. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E (2020) Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: Randomized controlled trial. JMIR mHealth uHealth 8(7):e15418. https://doi.org/10.2196/15418
22. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, Pulkki-Råback L, Baumeister H, Sander LB (2021) Digital interventions for the treatment of depression: A meta-analytic review. Psychol Bull 147(8):749–786. https://doi.org/10.1037/bul0000334

23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, Moher D (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71
24. Parsons EM, Ball T, Carl J, Henry AL, Emsley R, Miller CB, Wong EY, Coombs E, Collins D, Long LJ, Ster AC, Smits JAJ, Otto MW (2025) Digital cognitive behavioral treatment for generalized anxiety disorder: A randomized clinical trial. JAMA Netw Open 8(12):e2548884. https://doi.org/10.1001/jamanetworkopen.2025.48884
25. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, Earley C, Chapman J, Jell G, Sollesse S, Timulak L (2020) A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. npj Digit Med 3:85. https://doi.org/10.1038/s41746-020-0293-8
26. Salamanca-Sanabria A, Richards D, Timulak L, Connell S, Perilla MM, Parra-Villa Y, Castro-Camacho L (2020) A culturally adapted cognitive behavioral internet-delivered intervention for depressive symptoms: Randomized controlled trial. JMIR Mental Health 7(1):e13392. https://doi.org/10.2196/13392
27. Stiles-Shields C, Reyes KM, Lakhtakia T, Smith SR, Barnas OE, Gray EL, Krause C, Kruzan KP, Kwasny MJ, Mir Z, Panjwani S, Rothschild S, Sánchez-Johnsen L, Winquist N, Lattie EG, Allen NB, Reddy M, Mohr DC (2024) 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 24:828. https://doi.org/10.1186/s12888-024-06284-z
28. Terhorst Y, Kaiser T, Brakemeier E-L, Moshe I, Philippi P, Cuijpers P, Baumeister H, Sander LB (2024) Heterogeneity of treatment effects in internet- and mobile-based interventions for depression: A systematic review and meta-analysis. JAMA Netw Open 7(7):e2423241. https://doi.org/10.1001/jamanetworkopen.2024.23241

Figures

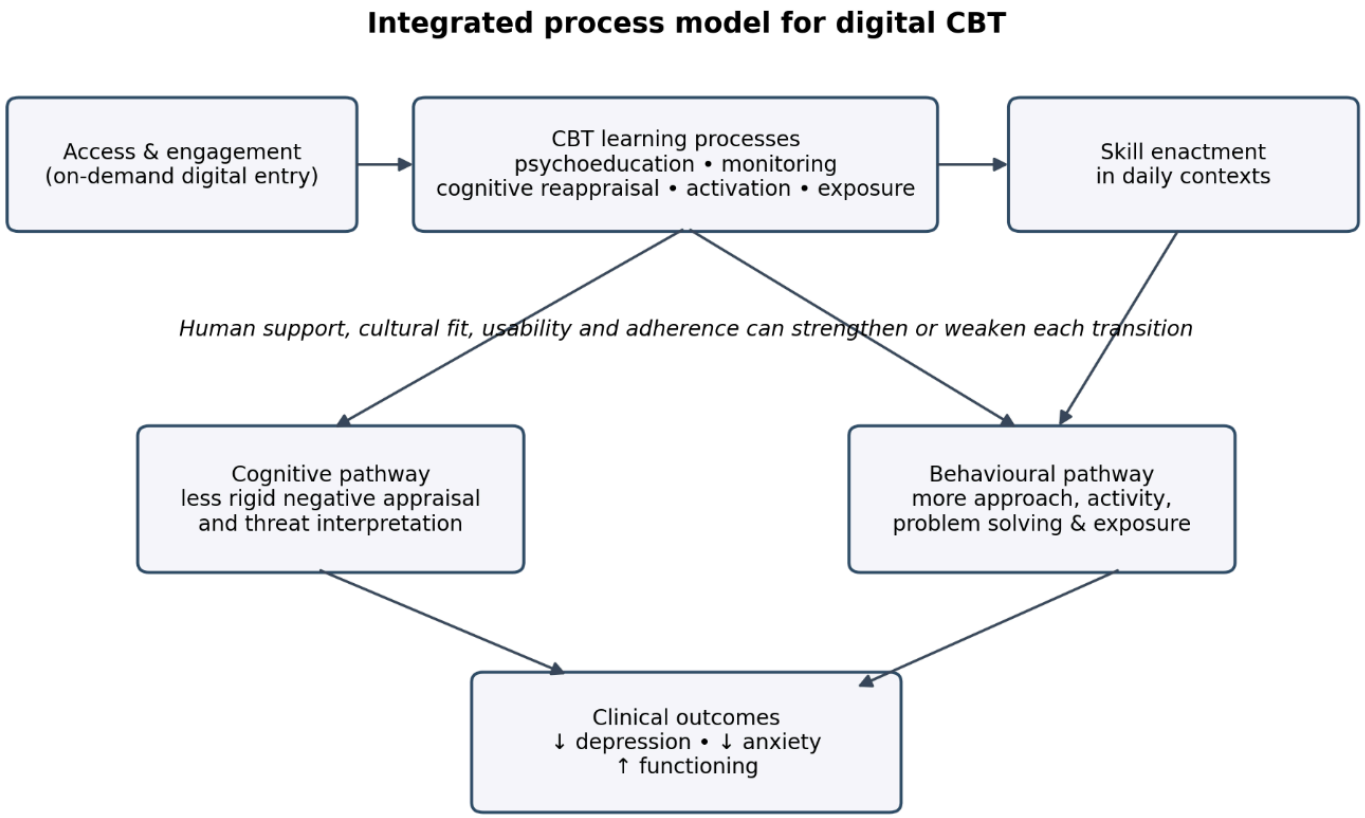


Figure 1

Integrated process model linking digital CBT delivery, skill acquisition and symptom change.

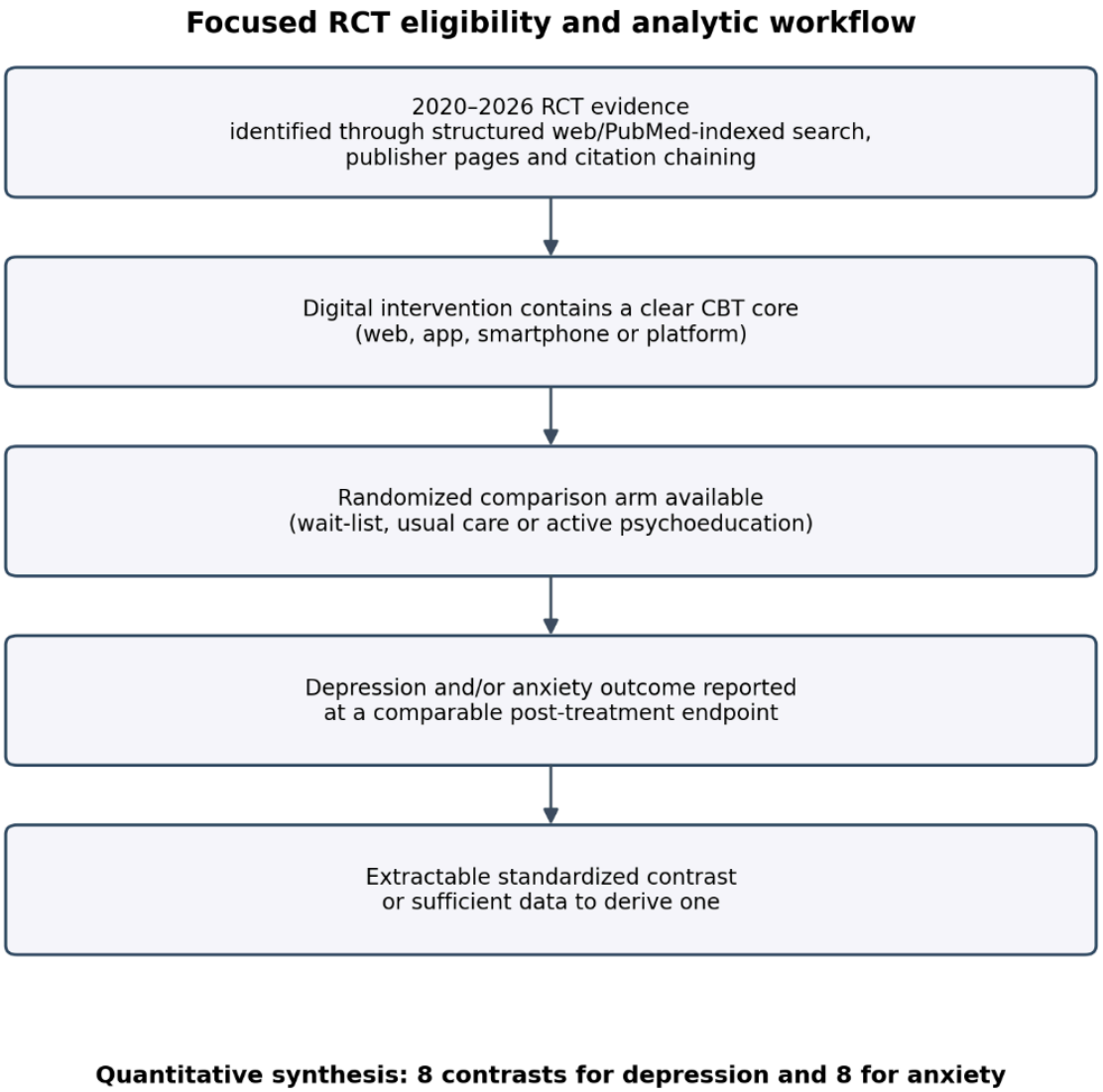


Figure 2

Focused RCT eligibility and analytic workflow used for the meta-analysis.

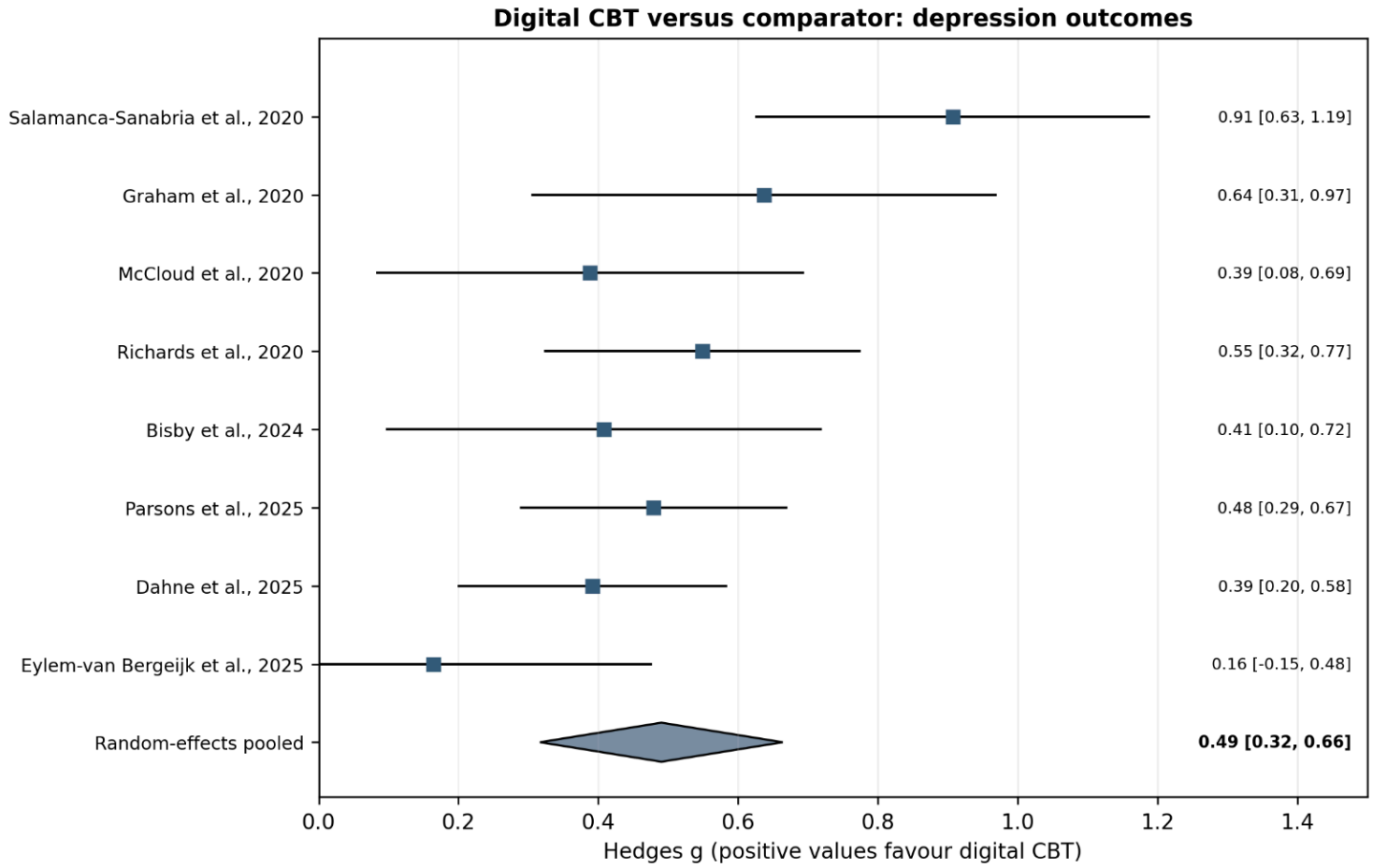


Figure 3

Random-effects forest plot for depressive symptoms. Positive Hedges g values favour digital CBT.

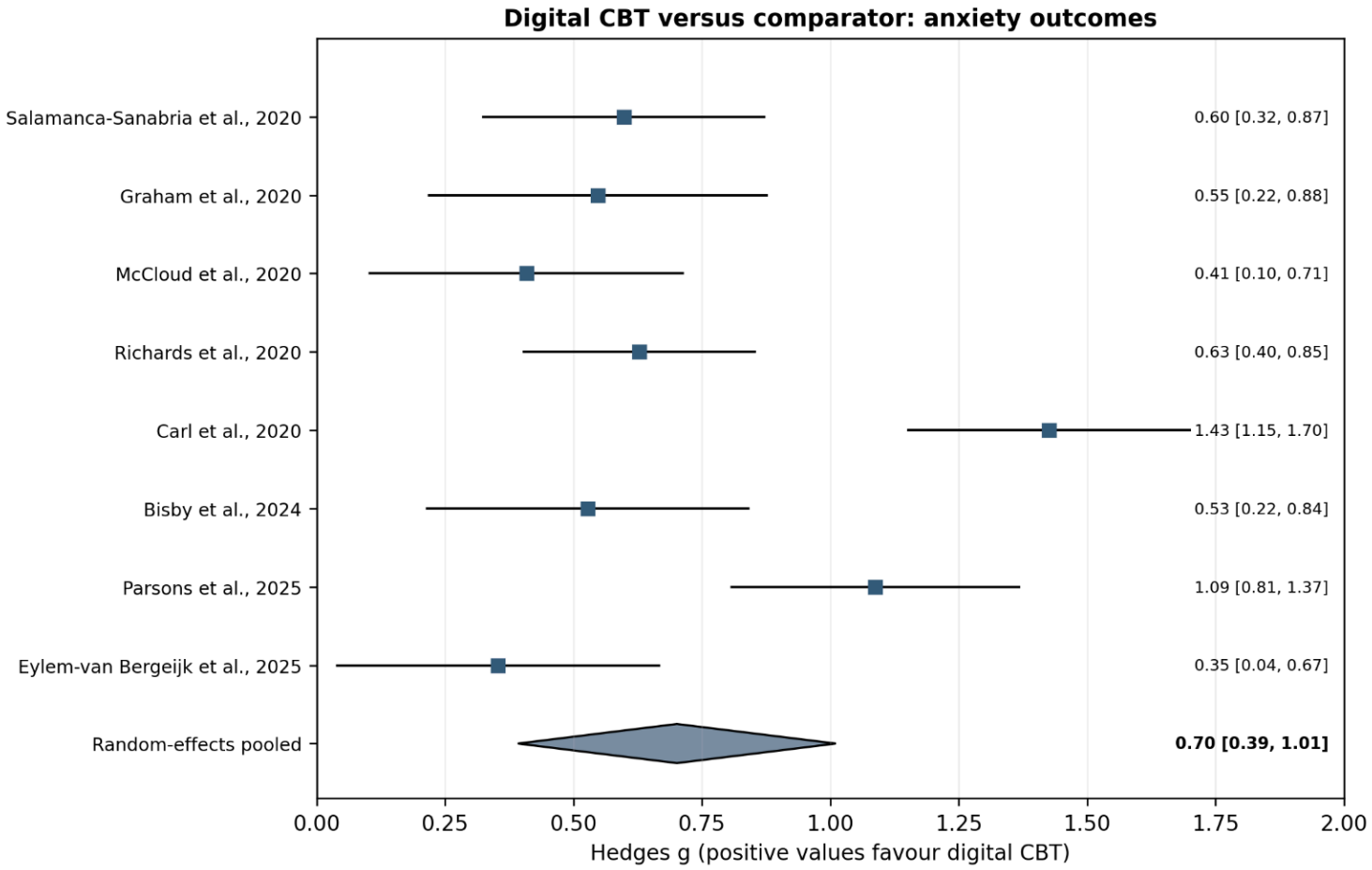


Figure 4

Random-effects forest plot for anxiety symptoms. Positive Hedges g values favour digital CBT.