

An umbrella review of digital psychological interventions for depressive symptoms in adolescents and young adults

Qingfeng Yang

Central China Normal University

Qinyu Sun

Central China Normal University

Jiawei Hu

Central China Normal University

Pei Yan

Central China Normal University

Shuai Chen

Central China Normal University

Zhihong Ren

ren@ccnu.edu.cn

Central China Normal University

Article

Keywords: Depressive symptoms, Digital psychological interventions, Adolescents, Young adults, Umbrella review, Meta-meta-analysis

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11129034/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Existing reviews of digital psychological interventions for youth depression differ in populations, comparators, and trial overlap. We conducted an umbrella review and meta-meta-analysis of interventions for people aged 10–25 years, searching seven databases through July 15, 2026. Quantitative synthesis was restricted to randomized evidence. Thirty-three reviews were included. The primary synthesis comprised nine review-level effects representing 62 independent randomized trial clusters. Relative to passive or nontherapeutic controls, digital psychological interventions reduced postintervention depressive symptoms (SMD = – 0.43, 95% CI – 0.52 to – 0.33; 95% prediction interval – 0.67 to – 0.19). Four active-control effects showed no clear difference (SMD = – 0.05, 95% CI – 0.33 to 0.22). Only two reviews remained after excluding critically low-quality reviews, precluding quality-restricted pooling. Certainty was moderate after downgrading for primary-trial risk of bias. Digital psychological interventions probably reduce postintervention depressive symptoms relative to passive or nontherapeutic controls; superiority to active treatments and sustained benefits remain uncertain.

Introduction

Depression is a common mental health problem in adolescence and young adulthood and contributes substantially to the burden of disease in this population^{1, 2}. This developmental period is marked by an increase in depressive symptoms and the onset of depressive disorders, alongside major transitions in education, relationships, and independent living². Depression can substantially disrupt interpersonal, educational, and social functioning^{2, 3}. Longitudinal evidence further links adolescent depression to adverse psychiatric and psychosocial outcomes in adulthood, although the strength of these associations varies across outcomes and may be influenced by earlier adversity, psychiatric comorbidity, and family or social circumstances^{4, 5}. Even when symptoms remain below the diagnostic threshold for a depressive disorder, subthreshold depression is associated with an increased risk of subsequent depression and suicide-related outcomes^{6, 7}. Group-based cognitive behavioral therapy delivered as an indicated preventive intervention can reduce depressive symptoms and the incidence of depressive disorders among young people with subthreshold symptoms, although these benefits may diminish over time⁸. Timely access to effective psychological support is therefore important throughout this developmental period.

Psychological interventions are an important component of depression prevention and treatment for young people. Clinical guidelines recommend structured approaches, including cognitive behavioral therapy and interpersonal psychotherapy, for adolescents with depression^{9, 10}. However, the availability of effective treatments does not ensure timely access: limited service capacity, costs, travel requirements, stigma, and concerns about confidentiality may all impede help-seeking and service use among young people¹¹. Digital psychological interventions offer a means of extending therapeutic support through internet platforms, computer programs, and mobile applications, sometimes incorporating messaging, video consultations, conversational agents, virtual reality, or therapeutic games^{12–14}. Digital delivery may reduce geographic and scheduling barriers while enabling self-paced participation, symptom monitoring, and feedback^{12, 15}. For digitally engaged young people who do not readily use conventional mental health services, these features may expand the reach and accessibility of psychological support^{13, 15}.

Over the past decade, systematic reviews and meta-analyses have examined an increasingly diverse range of digital psychological interventions relevant to depressive symptoms in young people. Earlier syntheses evaluated computerized psychological therapies for anxiety and depression in children and young people¹⁶ and internet- or computer-based cognitive behavioral interventions targeting anxiety or depression in participants up to 25 years of age¹⁷. Subsequent reviews examined additional digital formats and youth populations^{18–21}. Several depression-

related analyses reported effects favoring digital interventions, but the magnitude and precision of the estimates varied. Differences in participant characteristics, intervention formats, and comparison conditions make these findings difficult to interpret as a single, directly comparable body of evidence^{22, 23}. Confidence in some estimates is further constrained by the limited number and methodological quality of the underlying studies²⁴.

As the field has expanded, reviews have moved beyond internet- and computer-based cognitive behavioral therapy to examine mobile applications, mindfulness-based interventions, conversational agents, and therapeutic games^{25–30}. These interventions differ in therapeutic content, intensity, degree of human support, and mode of delivery. Their comparison conditions also vary, ranging from waitlists and no intervention to controls containing substantive therapeutic components. An observed benefit over a passive or nontherapeutic control therefore does not establish superiority over an active treatment. Network meta-analyses have provided additional evidence about relative effects within specific populations and treatment networks^{31–33}. However, because the populations, intervention nodes, comparators, and supporting evidence differ across these networks, their estimates should not be combined into a single cross-review treatment ranking.

Despite the growth of this literature, existing reviews do not consistently answer the more focused question of how digital psychological interventions affect depressive symptoms in people aged 10–25 years. Some reviews encompass broader age ranges or multiple mental health outcomes, making it difficult to isolate an estimate that simultaneously matches the population and outcome of interest. Reviews also differ in their treatment of randomized and nonrandomized studies, comparison conditions, and follow-up assessments^{19, 20, 22, 23, 30}. Variation in follow-up timing and effect definitions further limits cross-review conclusions about whether benefits persist beyond the end of treatment^{18, 20, 24}. Interpreting the accumulated evidence therefore requires explicit boundaries for population, outcome, study design, comparator, and assessment time point, rather than treating all reported effects as directly interchangeable.

An umbrella review can compare the scope, methods, and findings of existing meta-analyses within such boundaries, but the number of reviews should not be mistaken for the amount of independent primary evidence. The same randomized trial may contribute to several reviews; without identifying shared trials, a higher-level synthesis may overstate the breadth and independence of the evidence base³⁴. Methodological quality at the review level and risk of bias in the underlying trials must also be distinguished when judging the credibility of the evidence. Previous overviews have examined digital mental health interventions in children and young people^{12, 13}, as well as internet-based interventions across broader populations and mental health outcomes¹⁴. However, they do not directly resolve the combined questions of effect comparability, primary-trial overlap, and evidence certainty for depressive symptoms among people aged 10–25 years.

We therefore conducted an umbrella review and meta-meta-analysis of digital psychological interventions for depressive symptoms in adolescents and young adults aged 10–25 years. Our primary objective was to estimate the postintervention effect of these interventions relative to passive or nontherapeutic controls. We examined the robustness of this estimate to alternative effect definitions, age-eligibility rules, and evidence composition, while considering active-control and follow-up findings separately. We also identified primary trials shared across reviews, assessed the methodological quality of the included reviews, distinguished review-level quality from risk of bias in the underlying trials, and rated the certainty of evidence for the primary outcome. Together, these analyses were designed to clarify which conclusions are supported by the existing evidence, the populations and comparisons to which those conclusions apply, and the uncertainties that remain.

Results

Study selection

Database searches identified 13,227 records: 1,874 from PubMed, 3,722 from Embase, 3,234 from Web of Science Core Collection, 959 from CINAHL, 17 from APA PsycINFO, 109 from the Cochrane Database of Systematic Reviews, and 3,312 from Scopus. After removing 6,377 duplicates, 6,850 records underwent title and abstract screening; 6,743 were excluded. The full texts of the remaining 107 reports were obtained and assessed for eligibility.

The two reviewers initially agreed on 102 of the 107 full-text decisions (raw agreement, 95.33%; Cohen’s $\kappa = 0.89$). Five initial disagreements were resolved through discussion and adjudication involving J.H. Seventy-four full-text reports were excluded, primarily because they did not evaluate a digital psychological intervention ($n = 20$), did not meet population or age criteria ($n = 15$), did not permit separate extraction of an eligible digital-intervention effect ($n = 13$), contained no eligible original meta-analysis ($n = 8$), lacked a separately extractable depression outcome ($n = 7$), included only nonrandomized evidence ($n = 6$), lacked an eligible randomized between-group effect ($n = 2$), could not be sufficiently audited or mapped for quantitative analysis ($n = 1$), were not systematic reviews ($n = 1$), or had an ineligible publication type ($n = 1$). Thirty-three systematic reviews were included: 25 in the main evidence set and eight in the extended evidence set. Figure 1 shows the selection process; Supplementary Table S3 lists all excluded full-text reports and their primary reasons.

Characteristics of included systematic reviews

The 33 reviews were published between 2014 and 2026 and covered internet-based or computerized cognitive behavioral interventions, mobile applications, online mindfulness interventions, chatbots, digital therapeutic games, remote psychotherapy, and other digital psychological interventions. They differed in participant age and depressive symptom severity, theoretical basis and delivery of the intervention, level of human support, comparator, and outcome assessment time.

Table 1 summarizes core review characteristics; Supplementary Table S4 provides more detailed information on search scope, study design, numbers of primary studies and participants, statistical methods, and depression-specific effects. Twenty-five reviews were classified in the main evidence set and eight in the extended evidence set. For reviews with broader age scopes, eligibility was determined from the specific depression effect or intervention node: separately extractable target-age evidence was retained where available, whereas inseparable mixed-age effects were considered only under the prespecified 75% rule. Analysis-specific decisions are documented in Supplementary Table S2. Tracing primary studies across the 33 systematic reviews identified 164 independent RCTs meeting the present PICOS criteria, comprising 47,245 verified randomized participants. These totals describe primary randomized evidence meeting the present population, intervention, comparator, outcome, and design criteria; they are not the sum of overall study counts or sample sizes reported by the individual reviews.

Table 1
Characteristics of included systematic reviews

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
B. Chen et al. (2025) ³¹	Network meta-analysis	Main	Target age range: 10–25 years; samples with a mean age ≤ 25 years could include some older students	Internet-delivered psychotherapies	Waitlist, usual care, and active psychological interventions	27 reports; total <i>N</i> reported inconsistently as 3,345–3,451
Christ et al. (2020) ¹⁸	Pairwise meta-analysis	Main	Mean age: 12–25 years; elevated depressive symptoms or a depressive disorder	Computer-, internet-, or smartphone-delivered CBT	Passive and active controls	24 RCTs; passive-control depression pool: <i>k</i> = 20, <i>N</i> = 1,604; depression-targeted passive-control pool: <i>k</i> = 13, <i>N</i> = 1,162
Madrid-Cagigal et al. (2025) ³⁵	Pairwise meta-analysis	Main	Study mean ages: 18.53–25.00 years	Broad digital mental health interventions	Passive and active controls	34 studies in the narrative synthesis; 21 two-arm RCTs in the meta-analysis; depression: <i>k</i> = 16, <i>N</i> = 2,316
Ebert et al. (2015) ¹⁷	Pairwise meta-analysis	Main	Age range: 6–25 years; age-eligible depression analyses separately extractable	Computer- and internet-delivered CBT for anxiety and depression	Waitlist and active controls	13 RCTs; <i>N</i> = 796
Eilert et al. (2022) ¹⁹	Pairwise meta-analysis	Main	Children and young people	Internet-delivered psychological interventions	Mixed control conditions	23 studies in the systematic review; 16 in the meta-analysis; 21 comparisons, <i>N</i> = 1,985 (intervention, <i>n</i> = 977; control, <i>n</i> = 1,008)
Feng et al. (2025) ²⁶	Pairwise meta-analysis	Main	Average ages: 12–25 years	Artificial intelligence–driven conversational agents using	Waitlist, information-only, and active controls	14 articles reporting 15 trials; <i>N</i> = 1,974

Note. Main and extended evidence sets were defined a priori. Where analysis-specific *k* and *N* were available, these are reported alongside the overall review evidence. *k* = number of studies or comparisons; *N* = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
				natural language processing or machine learning		
Garrido et al. (2019) ³⁶	Mixed-methods review and pairwise meta-analysis	Main	Age range: 12–25 years	Broad digital mental health interventions	No-intervention and active controls	41 studies in the narrative synthesis; depression meta-analysis: 15 RCT comparisons, comprising 9 no-intervention and 6 active-control comparisons; <i>N</i> = NR
Harrer et al. (2019) ²²	Pairwise meta-analysis	Main	University students; eligibility required a mean age ≤ 24 years	Internet-based mental health interventions	Passive and active controls	48 RCTs; depression represented by 31 comparisons; <i>N</i> = NR
Hatoum et al. (2025) ³⁷	Pairwise meta-analysis	Main	Age range: 10–24 years; ≥80% of participants were age eligible	Digital single-session interventions	Passive and active controls; some uncontrolled evidence included narratively	22 articles describing 18 unique digital single-session interventions; 16 RCTs; total <i>N</i> = 9,353; depression follow-up meta-analysis: <i>k</i> = 9
Lin et al. (2026) ³⁸	Pairwise meta-analysis	Main	Young people without diagnosed mental disorders; samples predominantly comprised students and young adults	Mindfulness-based smartphone applications	Passive and active controls	19 RCTs in the review; 13 entered at least one meta-analysis; depression: <i>k</i> = 10, <i>N</i> = 1,543
Liu et al. (2024) ³⁹	Pairwise meta-analysis	Main	Study mean ages: 13.35–17.50 years	Self-guided or therapist-guided internet-delivered CBT	Waitlist, treatment as usual, and active controls	19 RCTs; postintervention depression: <i>k</i> = 16; preintervention <i>N</i> = 3,399
Note. Main and extended evidence sets were defined a priori. Where analysis-specific <i>k</i> and <i>N</i> were available, these are reported alongside the overall review evidence. <i>k</i> = number of studies or comparisons; <i>N</i> = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.						

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
Magwood et al. (2024) ²⁸	Pairwise meta-analysis	Main	Age range: 15–24 years	Mobile applications delivering mindfulness, CBT, combined, or other psychological interventions	Withheld and active controls	36 trials; <i>N</i> = 7,984
Noh et al. (2023) ⁴⁰	Pairwise meta-analysis	Main	Age range: 13–25 years	Website- and mobile-delivered psychological interventions, predominantly internet-delivered CBT	Passive and active controls	16 RCTs reported in 22 articles; <i>N</i> = NR
Noh and Kim (2023) ⁴¹	Pairwise meta-analysis	Main	Age range: 10–24 years	Online universal and selective prevention interventions	No intervention, waitlist, and active controls	19 RCTs; depression change-score analysis: <i>k</i> = 11, <i>N</i> = 2,687
Pennant et al. (2015) ¹⁶	Pairwise meta-analysis	Main	Age range: 5–25 years; depression analyses concerned young people	Computerized CBT, computerized interpretation-bias modification, computerized problem-solving therapy, and mobile applications	Nontherapeutic and active controls	27 RCTs; no single aggregate <i>N</i> reported; analysis-specific <i>k</i> and <i>N</i> available
Reangsing et al. (2023) ²⁹	Pairwise meta-analysis	Main	University students; overall mean age: 22.6 years (SD = 3.2); study mean ages: 18.2–31.1 years	Online mindfulness-based interventions	Passive and active controls	15 primary studies contributing 18 comparisons; <i>N</i> = 1,886
Sathe et al. (2026) ⁴²	Pairwise meta-analysis	Main	Age range: 13–25 years; the review also referred to a median-age criterion of 10–24 years	Mobile app-delivered CBT	Passive and active controls	10 RCTs; depression: <i>k</i> = 9; total review <i>N</i> = 1,896; main depression analysis <i>N</i> = 1,127
Note. Main and extended evidence sets were defined a priori. Where analysis-specific <i>k</i> and <i>N</i> were available, these are reported alongside the overall review evidence. <i>k</i> = number of studies or comparisons; <i>N</i> = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.						

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
Thabrew et al. (2018) ²⁴	Cochrane review and pairwise meta-analysis	Main	Included trials involved participants aged 10–18 years with long-term physical conditions	E-health psychological interventions	Waitlist, usual care, and attention controls	5 RCTs; $N = 463$
Tian et al. (2025) ⁴³	Pairwise meta-analysis	Main	Age range: 15–24 years; depressive and/or anxiety disorders	Telepsychiatry-delivered psychological interventions	Waitlist and active controls	26 studies, $N = 1,558$; 14 RCTs overall; depression versus waitlist: $k = 5$, $N = 293$
Townsend et al. (2022) ³⁰	Pairwise meta-analysis	Main	Age range: 12–25 years; the review also applied a mean-age criterion	Serious-game and gaming interventions	Passive and active controls	12 studies, including 7 RCTs and 5 non-RCTs; RCT effects separately extractable; RCT depression analysis: $N = 666$
Välimäki et al. (2017) ²⁰	Pairwise meta-analysis	Main	Included studies involved participants aged 11–24 years	Web-based psychological interventions	Passive and active controls	22 studies reported in 27 articles; 15 entered the meta-analysis, $N = 4,979$; postintervention depression: $k = 10$
Wang et al. (2023) ²³	Pairwise meta-analysis	Main	Adolescents and university students	Internet-based self-help interventions, including CBT, third-wave CBT, and problem-solving therapy	Mixed control conditions	25 comparisons; $N = 4,480$ (intervention, $n = 2,355$; control, $n = 2,125$)
Wang et al. (2026) ³³	Bayesian network meta-analysis	Main	Age range: 12–25 years	Four categories of digital health intervention, as defined by the source review	No intervention, treatment as usual, and active digital comparators	Depression network meta-analysis: 18 RCTs, $N = 5,821$, as reported by the source review; 17 depression-reporting RCTs identifiable in its study-

Note. Main and extended evidence sets were defined a priori. Where analysis-specific k and N were available, these are reported alongside the overall review evidence. k = number of studies or comparisons; N = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
						characteristics table
Wickersham et al. (2022) ²¹	Pairwise meta-analysis	Main	Age range: 11–19 years	Computerized CBT for depression prevention and treatment	Waitlist, usual care, and active controls	16 RCTs; 13 entered at least one meta-analysis; depression: $k = 10$; total review $N = 4,012$
Wu et al. (2023) ⁴⁴	Pairwise meta-analysis	Main	Age range: 11–19 years	Internet- and computer-delivered CBT for adolescent depression	Waitlist, usual care, and active controls	18 RCTs; $N = 1,683$
Bevilacqua et al. (2024) ⁴⁵	Pairwise meta-analysis	Extended	Review eligibility required age ≤ 18 years, although one study included participants aged 15–25 years	Technology-assisted CBT	Face-to-face CBT	10 RCTs overall; depression: $k = 7$; $N = \text{NR}$
Bryant et al. (2024) ⁴⁶	Pairwise meta-analysis	Extended	Pediatric samples; review eligibility required age < 18 years	Gamified digital mental health interventions	Passive and active controls	27 RCTs; $N = 2,911$; depression: $k = 9$, $N = 1,332$
T. H. Chen et al. (2025) ²⁵	Pairwise meta-analysis	Extended	Overall age range: 14.7–35.3 years; eligible university subgroup mean ages: 18.78–22.90 years	CBT-informed conversational agents	Passive and active controls	10 RCTs overall; depression: $k = 9$; age-eligible subgroup: $k = 3$, $N = 205$
Grist et al. (2019) ⁴⁷	Pairwise meta-analysis	Extended	Age range: 6–18 years; depression-targeted subgroup separately extractable	Computer-, internet-, app-, and virtual reality-delivered psychological interventions	Waitlist, treatment as usual, and active controls	34 RCTs; $N = 3,113$; depression-targeted subgroup: 13 comparisons
Leech et al. (2021) ²⁷	Pairwise meta-analysis	Extended	Age range: 10–35 years; project audit confirmed	Standalone smartphone applications	Passive and active controls	11 independent RCTs; total $N = 1,706$;
Note. Main and extended evidence sets were defined a priori. Where analysis-specific k and N were available, these are reported alongside the overall review evidence. k = number of studies or comparisons; N = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.						

Review	Review design	Evidence set	Population and age eligibility	Digital intervention scope	Comparator scope	Included evidence (overall; eligible depression)
			that $\geq 75\%$ of the evidence was age eligible			depression: $k = 8$, $N = 862$
Li et al. (2021) ³²	Bayesian network meta-analysis	Extended	Review age range: 7–18 years; participants in all three eligible computerized CBT trials were aged 11–16 years	Broad psychosocial prevention interventions with a separable computerized CBT node	Waitlist and other psychosocial intervention nodes	27 RCTs; $N = 5,976$; eligible computerized CBT node: $k = 3$
Takizawa et al. (2026) ⁴⁸	Pairwise meta-analysis	Extended	Overall age range: 5–19 years; the depression pool predominantly comprised participants aged 10–19 years	Universal digital mental health interventions	No intervention, usual practice, and active controls	29 RCTs; $N = 22,705$; depression: $k = 15$
Ye et al. (2014) ⁴⁹	Pairwise meta-analysis	Extended	Children, adolescents, and young adults; the depression waitlist pool was audited against the project age criterion	Internet-based psychological interventions for anxiety and/or depression	Waitlist and face-to-face psychological interventions	Depression versus waitlist: $k = 7$, $N = 490$; internet versus face-to-face intervention: $k = 2$, $N = 66$
Note. Main and extended evidence sets were defined a priori. Where analysis-specific k and N were available, these are reported alongside the overall review evidence. k = number of studies or comparisons; N = total sample size; CBT = cognitive behavioral therapy; NR = not reported; RCT = randomized controlled trial. Detailed review characteristics and age-eligibility assessments are provided in Supplementary Table S4.						

Size of the primary RCT evidence base and overlap

We linked RCTs reported in review articles, supplements, forest plots, and study-characteristics tables, merging primary-outcome, follow-up, subgroup, and process-evaluation reports from the same randomized trial into one independent RCT cluster.

The nine reviews included in the strict primary analysis contained 83 review–RCT occurrences corresponding to 62 independent RCT clusters. The model-specific corrected covered area was 4.23%, indicating low overall overlap in the strict primary analysis. Nevertheless, some reviews shared several trials, particularly broader reviews and earlier topic-specific reviews.

RCT membership for the strict primary analysis is documented in Supplementary Table S5, and pairwise sharing of independent RCT clusters is shown in Supplementary Figure S1. Overlap was used to characterize dependence in the

underlying evidence and contextualize the pooled result; it did not alter the source-reported review-level effects entering the meta-meta-analysis.

Methodological quality of included reviews

Two reviewers independently assessed all 33 reviews using AMSTAR 2. Their initial judgments agreed on 426 of 528 paired items (raw agreement, 80.68%; Cohen's $\kappa = 0.63$), indicating substantial overall item-level agreement. Disagreements centered on prospective protocols, search comprehensiveness, justification of eligible study designs, and adequacy of study descriptions. All were discussed after rechecking review articles, supplements, and registration records, with J.H. adjudicating where necessary; final ratings reflected consensus.

No review was rated high quality by AMSTAR 2. One (3.03%) was rated moderate, five (15.15%) low, and 27 (81.82%) critically low. Common critical weaknesses were the absence of a study-by-study list of full-text exclusions with reasons, searches that did not meet all AMSTAR 2 criteria for comprehensiveness, insufficiently verifiable a priori protocols, and inadequate assessment or discussion of publication bias. Common noncritical weaknesses included failure to report funding sources for primary studies, insufficient justification for eligible study designs, and incomplete reporting of study characteristics.

The item-level distribution is presented in Supplementary Figure S2; all 16 item judgments and overall confidence ratings for each review are given in Supplementary Table S6. Because overall AMSTAR 2 ratings depend on critical weaknesses rather than a simple item sum, we did not calculate a numerical quality score or automatically exclude critically low-quality reviews from the primary analysis.

Overall effect of digital psychological interventions on depressive symptoms

The strict primary analysis included nine review-level effects from nine systematic reviews, representing 62 independent RCT clusters. Source-reported Hedges' g , Cohen's d , or SMD estimates were harmonized in direction and entered the meta-meta-analysis as standardized mean differences (SMDs), with negative values favoring digital psychological intervention over passive or nontherapeutic control.

The random-effects meta-meta-analysis showed lower depressive symptom severity at postintervention for digital psychological interventions than for passive or nontherapeutic controls (SMD = -0.43 , 95% CI -0.52 to -0.33 , $p < 0.001$; Fig. 2). Review-specific estimates and their source locations are documented in Supplementary Table S2.

Review-level estimates had a consistent direction and low-to-moderate statistical heterogeneity ($Q = 12.36$, $df = 8$, $p = 0.136$; $I^2 = 35.27\%$; $\tau^2 = 0.0087$). The 95% prediction interval (-0.67 to -0.19) did not cross zero, suggesting that the true effect would be expected to favor digital psychological intervention in future evidence settings with clinical and methodological characteristics similar to those represented in the strict primary analysis. The CCA was 4.23%, indicating low overall overlap of underlying RCTs across reviews.

Supporting and sensitivity analyses

Change-score sensitivity analysis

Adding the standardized between-group change-score effect reported by Noh and Kim⁴¹ to the strict primary analysis produced a model with 10 review-level effects. Digital psychological interventions remained associated with reduced

depressive symptoms at postintervention (SMD = - 0.39, 95% CI - 0.52 to - 0.26, $p < 0.001$). Heterogeneity was greater than in the strict primary analysis ($Q = 36.00$, $df = 9$, $p < 0.001$; $I^2 = 75.00\%$; $\tau^2 = 0.0232$), although the 95% prediction interval remained below zero (- 0.76 to - 0.02). The CCA was 3.81%, indicating low primary-RCT overlap. Thus, the effect direction and statistical inference were retained when the change-score effect was included, but between-review variation increased substantially (Table 2; Supplementary Figure S3).

Leave-one-review-out analysis

Across leave-one-review-out models, pooled effects ranged from - 0.48 to - 0.41; all 95% confidence and prediction intervals remained below zero, and I^2 ranged from 0.00% to 43.32%. Omitting Harrer et al.²² yielded an SMD of - 0.48 (95% CI - 0.55 to - 0.40) and reduced I^2 to 0.00%, indicating that this review was a major contributor to heterogeneity. Omitting Ebert et al.¹⁷ or Christ et al.¹⁸ produced the smallest absolute pooled effects, both approximately - 0.41, without changing direction or inference. No single review drove the primary result (Supplementary Table S7; Supplementary Figure S4).

Extended-evidence sensitivity analysis

The extended-evidence sensitivity analysis included 11 review-level effects to test robustness to broader age-composition or effect-extraction criteria. The pooled estimate still favored digital psychological intervention (SMD = - 0.37, 95% CI - 0.49 to - 0.24, $p < 0.001$), but heterogeneity was high ($I^2 = 86.00\%$; $\tau^2 = 0.0270$), and the 95% prediction interval crossed zero (- 0.76 to 0.03). The CCA was 3.29%, indicating low overall overlap; overlap alone was therefore unlikely to account fully for the increased heterogeneity. Relative to the strict primary analysis, the average effect was slightly smaller and estimates were more dispersed, consistent with greater variation in true effects when eligibility boundaries were broadened (Table 2). Primary-study mapping coverage and overlap for this model are documented in Supplementary Table S8.

Active-control subgroup analysis

The active-control subgroup included four review-level effects. There was no clear difference between digital psychological interventions and active controls (SMD = - 0.05, 95% CI - 0.33 to 0.22, $p = 0.580$); heterogeneity was moderate ($Q(3) = 7.03$, $p = 0.071$; $I^2 = 57.40\%$; $\tau^2 = 0.0070$). A prediction interval was not calculated because this model contained only four effects. The CCA was 1.67%, indicating low primary-RCT overlap. Thus, the advantage over passive or nontherapeutic controls should not be extrapolated to superiority over interventions with substantive therapeutic components (Table 2; Supplementary Figure S5).

Methodological-quality-restricted analysis

Among the nine reviews in the strict primary analysis, AMSTAR 2 rated only Thabrew et al.²⁴ moderate and Ebert et al.¹⁷ low; the remaining seven were critically low. Excluding critically low-quality reviews left only two review-level effects, too few to estimate between-review variance and heterogeneity reliably. We therefore did not perform a quality-restricted quantitative re-synthesis. Both retained effects favored digital psychological intervention, although the confidence interval for Thabrew et al.²⁴ crossed zero. These findings were interpreted descriptively and do not establish whether the pooled primary effect would persist in a sufficiently large set of higher-quality reviews (Supplementary Table S9).

Follow-up and network meta-analysis evidence

Follow-up effects differed substantially in timing, comparator, and reporting, precluding a coherent meta-meta-analytic model corresponding to the strict primary analysis. We therefore synthesized these results narratively. Some reviews suggested that benefits persisted into short- or medium-term follow-up, but effects were generally attenuated relative to postintervention and not all comparisons were statistically significant. Given the small and temporally inconsistent follow-up evidence base, these findings are supportive rather than sufficient to establish the duration of benefit (Supplementary Figure S6).

Two reviews contributed the network meta-analysis estimates displayed in Supplementary Figure S7. B. Chen et al.³¹ compared internet-based psychotherapies, whereas Li et al.³² reported network estimates for computerized cognitive behavioral therapy against different controls. Although the estimates generally favored digital intervention, they came from networks with different structures, comparison nodes, and Bayesian models, and effects within each network were correlated. We therefore retained source-reported network estimates as supplementary descriptions of differences by intervention and comparator, without pooling across networks or including them in the strict primary or supporting meta-meta-analyses (Supplementary Figure S7).

Table 2
Primary, supporting, and sensitivity analyses

Analysis	k	SMD (95% CI)	I ² (%)	τ ²	95% prediction interval	CCA (%)	Analytical role
Strict primary analysis	9	−0.43 (−0.52 to −0.33)	35.27	0.0087	−0.67 to −0.19	4.23	Primary
Change-score sensitivity analysis	10	−0.39 (−0.52 to −0.26)	75.00	0.0232	−0.76 to −0.02	3.81	Sensitivity
Extended evidence-set sensitivity analysis	11	−0.37 (−0.49 to −0.24)	86.00	0.0270	−0.76 to 0.03	3.29	Sensitivity
Active-control subgroup	4	−0.05 (−0.33 to 0.22)	57.40	0.0070	Not calculated (k < 5)	1.67	Supporting subgroup
Note. Negative SMD values favour digital psychological interventions. Prediction intervals were prespecified for models containing at least five review-level effects. CI = confidence interval; CCA = corrected covered area.							

Small-study effects and publication bias

The strict primary analysis included only nine review-level effects, below the usual minimum for formal funnel-plot asymmetry testing. Moreover, the reviews shared some primary RCTs, so review-level effects were not fully independent. We therefore did not perform Egger’s test or another formal asymmetry test for this model. Publication-bias assessments varied across source reviews: some reported funnel plots, Egger tests, or trim-and-fill analyses, with isolated findings suggesting possible small-study effects, although corrected effect directions generally remained unchanged; other reviews found no clear evidence or did not test because too few RCTs were available. Overall, publication bias was not strongly suspected but cannot be excluded.

Certainty of evidence

GRADE was used to assess certainty for the strict primary outcome. Randomized evidence started at high certainty. We downgraded one level for risk of bias because some primary RCTs had missing outcome data, relied on self-reported outcomes, showed deviations from intended interventions, or raised concerns about selective reporting;

some source reviews also reported attenuation when analyses were restricted to lower-risk trials. We did not downgrade for inconsistency because all nine review-level effects favored digital intervention, statistical heterogeneity was low to moderate, and the 95% prediction interval did not cross zero. The strict primary analysis directly matched the prespecified population, digital intervention, passive or nontherapeutic comparator, depressive-symptom outcome, and postintervention time point, so indirectness did not warrant downgrading. Neither the pooled 95% confidence interval nor the 95% prediction interval crossed zero, and the evidence comprised 62 independent RCT clusters; therefore, we did not downgrade for imprecision. Publication bias could not be ruled out but was not strongly suspected and did not prompt downgrading. Overall certainty was moderate, indicating that digital psychological interventions probably reduce depressive symptoms at postintervention in young people aged 10–25 years (Table 3; Supplementary Table S10).

The two reviewers initially agreed on 80.00% of judgments across the five GRADE domains. The disagreement concerned whether to downgrade for publication bias; after discussion and third-reviewer adjudication, the final decision was not to downgrade.

Table 3
Summary of Findings and certainty assessment

Outcome	Effect	Evidence base	GRADE	Interpretation
Post-intervention depressive symptoms	SMD – 0.43 (95% CI – 0.52 to – 0.33); 95% PI – 0.67 to – 0.19	9 reviews; 62 independent RCT clusters; CCA 4.23%	Moderate	Digital psychological interventions probably reduce depressive symptoms; certainty was downgraded by one level for risk of bias in the primary RCTs.
<i>Note. Formal tests of funnel-plot asymmetry were not performed because the strict primary analysis included fewer than 10 review-level effects and the reviews shared primary studies. CI = confidence interval; PI = prediction interval; CCA = corrected covered area.</i>				

Discussion

To our knowledge, this is the first umbrella review and meta-meta-analysis to focus on depressive outcomes in adolescents and young adults aged 10–25 years while synthesizing review-level evidence within a consistent framework for age, study design, comparator, and outcome. We included 33 systematic reviews, of which nine contributed comparable effects representing 62 independent RCT clusters to the strict primary analysis. Digital psychological interventions probably reduced depressive symptoms at postintervention relative to passive or nontherapeutic controls (SMD = – 0.43, 95% CI – 0.52 to – 0.33), and the 95% prediction interval remained entirely in the favorable direction (– 0.67 to – 0.19). This conclusion was supported by moderate-certainty evidence, indicating reasonable confidence in the direction of effect while leaving open the possibility that risk of bias in the underlying RCTs may affect its true magnitude. The finding should therefore be interpreted within the population, intervention, comparator, and assessment conditions represented in the strict primary analysis rather than generalized to all digital psychological interventions or all young people.

The direction of the primary result is consistent with previous reviews of digital psychological interventions for children, adolescents, and young adults. Ebert et al.¹⁷ reported improvements in depressive symptoms following internet- and computer-based cognitive behavioral interventions for young people, whereas Christ et al.¹⁸ found a similarly favorable direction in randomized evidence on internet interventions for children and adolescents. Pennant et al.¹⁶ and Thabrew et al.²⁴ also reported evidence suggesting benefits from technology-delivered psychological interventions, although the interpretation of their estimates was constrained by limited study numbers, intervention

heterogeneity, and the quality of the underlying trials. The present strict primary analysis improved comparability by restricting synthesis to randomized evidence, postintervention depressive symptoms, and passive or nontherapeutic controls. Nevertheless, the pooled SMD remains an average standardized difference across diverse interventions and depression scales and should not be equated with clinical remission, diagnostic recovery, or improvement that an individual would necessarily perceive as clinically important.

Comparator type was central to the clinical interpretation of the findings. The benefit observed against passive or nontherapeutic controls did not translate into a clear advantage over active controls: the active-control estimate was close to zero and imprecise (SMD = - 0.05, 95% CI - 0.33 to 0.22). Harrer et al.²² similarly showed that digital-intervention effects varied by comparator and were generally larger against passive than active controls, while the different comparison conditions included by Pennant et al.¹⁶ and Wickersham et al.²¹ further illustrate the importance of control-group intensity. Because active controls contain substantive therapeutic components, a smaller between-group difference does not imply that digital psychological interventions are ineffective. Rather, the current evidence supports a benefit relative to receiving no substantive treatment but has not established superiority over other interventions containing therapeutic components. Given that only four review-level effects contributed to the active-control analysis and the confidence interval crossed zero, this finding reflects limited comparative evidence rather than demonstrated equivalence.

The primary conclusion was stable when change-score effects were included and when individual reviews were omitted, whereas broadening the evidence set substantially increased heterogeneity. Adding a standardized between-group change-score effect retained a favorable pooled estimate (SMD = - 0.39, 95% CI - 0.52 to - 0.26), and no leave-one-review-out analysis altered the direction or principal inference. The primary result was therefore not dependent on any single review or entirely contingent on the use of postintervention rather than change-score effects. However, heterogeneity increased in the change-score model, potentially reflecting differences in study composition and the construction of change-score estimates. The extended evidence set also produced a favorable average effect (SMD = - 0.37, 95% CI - 0.49 to - 0.24), but heterogeneity increased further and the prediction interval crossed zero. Broadening the eligibility boundary therefore increased the range of evidence considered while reducing the consistency expected across comparable future settings. The extended analysis is most appropriately interpreted as a test of sensitivity to eligibility boundaries rather than as a replacement for the strict primary analysis.

The age range of 10–25 years was selected to span the developmental continuum from adolescence to young adulthood, consistent with a previous systematic review of e-health interventions for children and young people⁵⁰. However, the source reviews generally reported study-level mean ages or age ranges rather than individual participant data. The extended-evidence analysis can therefore evaluate the sensitivity of the primary conclusion to the age-eligibility rule but cannot establish whether age continuously modifies intervention effects or reliably compare outcomes across early adolescence, later adolescence, and young adulthood. Primary RCTs should report age-stratified results where appropriate, while individual participant data meta-analysis may be better suited to evaluating age-related effect modification when sufficient data become available.

Follow-up and network meta-analysis evidence provided additional context but did not support firm conclusions about the duration of benefit or a common ranking of digital interventions. Follow-up assessments differed in timing, comparator, and effect definition. Christ et al.¹⁸ reported follow-up findings for some internet interventions, but the number of contributing studies and duration of follow-up varied; follow-up evidence in Thabrew et al.²⁴ and Välimäki et al.²⁰ was also substantially sparser than postintervention evidence. The primary conclusion should therefore be restricted to depressive symptoms at postintervention and should not be interpreted as establishing sustained benefit. B. Chen et al.³¹ compared internet-based psychotherapies, Li et al.³² estimated the effects of computerized

cognitive behavioral therapy against several controls, and Wang et al.³³ compared digital intervention types within its own evidence network. Differences in populations, intervention and comparator nodes, network structures, and supporting trials precluded pooling these estimates or deriving a common cross-review treatment ranking.

The moderate GRADE certainty assigned to the primary outcome should be distinguished from the generally low methodological confidence in the included reviews. None of the 33 reviews received a high overall confidence rating under AMSTAR 2; one was rated moderate, five low, and 27 critically low. Common weaknesses included the absence of a sufficiently verifiable a priori protocol, incomplete reporting of excluded full-text studies, inadequate searches, and insufficient assessment or discussion of publication bias. The increasing number of reviews in this field has therefore not been matched by equivalent methodological rigor. Nevertheless, a critically low AMSTAR 2 rating does not demonstrate that a digital intervention is ineffective or that every RCT included in the review has serious bias. AMSTAR 2 concerns confidence in the conduct and reporting of a review, whereas the GRADE risk-of-bias domain concerns the underlying trials supporting the target effect.

We therefore did not automatically exclude reviews according to their AMSTAR 2 rating but prespecified a quality-restricted analysis. After exclusion of critically low-confidence reviews, only Thabrew et al.²⁴ and Ebert et al.¹⁷ remained. Two review-level effects were insufficient for reliable estimation of between-review variance and heterogeneity, and we therefore did not undertake a potentially misleading quantitative re-synthesis. This finding does not show that the primary effect disappears among reviews with greater methodological confidence; rather, too few reviews not rated critically low were available to verify the primary result independently. GRADE certainty was downgraded by one level because of risk of bias in the underlying RCTs, including missing outcome data, reliance on self-reported outcomes, deviations from intended interventions, and concerns about selective reporting. More rigorous RCTs could therefore revise the estimated magnitude, although the current evidence consistently favors a reduction in depressive symptoms at postintervention.

Mapping of overlapping RCTs indicated limited overall duplication in the strict primary model: 83 review–RCT occurrences represented 62 independent RCT clusters, corresponding to a CCA of 4.23%. Overlap was also low in the change-score, extended-evidence, and active-control models. However, a low CCA does not eliminate statistical dependence among review-level effects because it summarizes the overall overlap structure rather than the contribution of each shared RCT to each review estimate. Identification and reporting of shared primary studies make the evidence structure more transparent but do not by themselves remove this dependence. Future umbrella reviews should report the primary-study membership of each synthesis and, where the necessary information is available, consider methods that accommodate correlated review-level effects.

Digital psychological interventions may provide an additional route for delivering depression-related psychological support to adolescents and young adults, but the present evidence does not support their use as a universal replacement for other effective treatments. The benefit observed relative to passive or nontherapeutic controls suggests that these interventions may be useful when no substantive treatment is otherwise provided or as a complement to existing services. Digital delivery may reduce some temporal and geographic barriers and offer an additional entry point for young people who cannot promptly access conventional services¹². However, the active-control analysis showed no clear additional benefit, and digital interventions cannot yet be considered superior to interventions containing substantive therapeutic components. Clinical decisions should therefore consider therapeutic content, target population, degree of human support, and individual need rather than digital delivery alone. Because crisis management and safety in high-risk populations were not evaluated, the present findings do not justify replacing professional assessment and ongoing monitoring with digital intervention for young people with severe symptoms or elevated safety risk.

Future RCTs should determine the incremental benefit of digital psychological interventions against clearly defined comparison conditions and establish whether benefits persist beyond treatment. Waitlist and no-intervention controls address whether an intervention produces benefit relative to receiving no substantive treatment, whereas active controls provide a more demanding assessment against another intervention. Combining these conditions obscures the distinction between benefit over minimal treatment and superiority over an alternative treatment. Trials should therefore describe the services actually received by control participants and report passive- and active-control comparisons separately. Because the current evidence is centered on postintervention depression-scale scores, future studies should include clinically meaningful follow-up and assess clinical recovery, everyday functioning, engagement, and safety alongside symptom change. These outcomes are necessary to determine whether digital psychological interventions can have a sustained role in routine youth mental health care.

Greater consistency in intervention descriptions and outcome reporting is also needed to improve comparability. Digital psychological interventions differ not only in delivery technology but also in therapeutic basis, automation, intervention intensity, and human support, while source reviews have applied inconsistent classifications. Dietvorst et al.³⁷ similarly identified variation in therapeutic principles, delivery modes, personnel involvement, and interaction design across youth e-health interventions. Future trials should clearly specify their core therapeutic components, intervention intensity, degree of human support, and actual intervention use, thereby enabling more informative analyses of which approaches benefit which young people. Rather than repeatedly producing broad and highly overlapping reviews, the field now requires high-quality RCTs with explicit comparative questions, adequate follow-up, and transparent reporting. Individual participant data meta-analysis could then examine modification by age or symptom severity, while network meta-analysis using consistently defined nodes could compare relative effects across intervention types.

This review has several strengths. We prespecified the target age range, study design, comparator, outcome, and assessment time and considered the strict primary analysis separately from change-score, extended-evidence, active-control, and follow-up evidence, thereby avoiding direct pooling of effects with different clinical meanings. Because primary-study overlap is an important methodological concern in umbrella reviews⁵¹, we mapped the included reviews to their underlying RCTs and calculated CCA from the studies contributing to each model³⁴. Random-effects models, Knapp–Hartung adjustment, and prediction intervals were used to represent uncertainty in the pooled estimates. We assessed methodological confidence in the included reviews using AMSTAR 2⁵² and certainty in the primary outcome using GRADE⁵³. Together, these procedures made the robustness, credibility, and boundaries of applicability of the primary finding more transparent.

Several limitations should inform interpretation. First, the meta-meta-analysis relied on pooled effects reported by the source reviews and therefore depended on their selection of primary studies, data processing, and statistical methods. Although we mapped RCT membership for the prespecified models and checked source information where necessary, we could not independently reconstruct every analysis from participant-level or trial-level outcome data. Second, some reviews shared primary RCTs. A low CCA indicates limited overall duplication but does not eliminate dependence among review-level effects or account for the influence of each shared RCT on the source estimates³⁴. Finally, active-control and follow-up evidence was limited, precluding firm conclusions about superiority over other effective interventions or the duration of benefit. The strongest inference is therefore restricted to depressive symptoms at postintervention relative to passive or nontherapeutic controls.

Methods

Study design and protocol registration

We conducted an umbrella review and meta-meta-analysis and reported the study in accordance with the PRIOR statement and PRISMA 2020 guidelines^{54, 55}. The protocol was registered in PROSPERO on August 5, 2026 (CRD420261445516). At registration, database searching, deduplication, and title and abstract screening had been completed, and full-text assessment was underway; data extraction and evidence synthesis had not begun. During protocol development, we searched PROSPERO and OSF Registries on July 14, 2026 and identified no registered review that substantially duplicated the present study. Relative to the registered protocol, included reviews were appraised using AMSTAR 2, but a separate ROBIS assessment was not undertaken because of overlap between the tools. We additionally applied GRADE to the strict primary outcome after mapping the underlying randomized controlled trials (RCTs) and tracing their risk-of-bias information.

Eligibility criteria

We included systematic reviews containing a meta-analysis or network meta-analysis relevant to adolescents and young adults aged 10–25 years. This age range was selected to span the developmental continuum from adolescence to young adulthood and is consistent with a previous systematic review of e-health interventions for young people⁵⁰. Eligible reviews evaluated interventions delivered primarily through digital technology, with an identifiable psychological therapeutic component, and reported depressive symptoms measured using validated scales. Eligible modalities included internet- or web-based interventions, computerized psychotherapy, mobile applications, digital mindfulness interventions, therapeutic games, conversational agents, text-message interventions, virtual reality, and other technology-assisted psychological interventions. Digital tools used solely for screening, monitoring, administration, or general health information, without a psychological intervention component, were excluded.

Age eligibility was determined for the depression effect relevant to each analysis rather than from the overall age range of the review. We retained separately extractable effects for participants aged 10–25 years, including eligible intervention nodes in network meta-analyses whose contributing trials met the target-age criterion. A mixed-age effect without a separately extractable target-age result was eligible for the extended evidence set only when at least 75% of the participants contributing to that effect could be reliably verified as aged 10–25 years. Otherwise, it was excluded from quantitative synthesis. Classification of a review in the main or extended evidence set did not imply that all effects reported by that review were eligible.

The 75% threshold was a prespecified pragmatic rule informed by majority-sample criteria used in previous reviews^{56, 57}. Because our unit of analysis was a review-reported pooled effect, we applied the threshold to participants contributing to the relevant depression estimate rather than to the review's total sample.

Quantitative synthesis was restricted to individually or cluster-randomized controlled trials. Reviews containing randomized and nonrandomized studies were eligible only when an RCT-specific result could be extracted separately. Eligible comparators included waitlist, no intervention, usual care, minimal information or psychoeducation, attention controls or digital placebos, and active psychological interventions. Passive or nontherapeutic controls were separated a priori from controls containing substantive therapeutic components. Usual care, psychoeducation, and attention controls were classified according to their reported content rather than their labels alone.

The primary outcome was depressive symptom severity at postintervention. Follow-up findings were classified into prespecified windows of 1–3, > 3–6, and > 6–12 months after intervention completion. For reviews reporting multiple

mental health outcomes, only separately identifiable depression results were eligible.

We restricted eligible publications to peer-reviewed systematic reviews and meta-analyses published in English and available in full text, although no language restriction was applied during database searching. Narrative reviews without quantitative synthesis, scoping reviews, evidence maps, protocols, conference abstracts, dissertations, correspondence, commentaries, and other non-peer-reviewed reports were excluded.

Search strategy

We searched PubMed, Embase, Web of Science Core Collection, CINAHL, APA PsycINFO, the Cochrane Database of Systematic Reviews, and Scopus from database inception through July 15, 2026. Search strategies were developed from the prespecified PICOS framework and combined controlled vocabulary and free-text terms for depression, adolescence and young adulthood, digital delivery, psychological intervention, and systematic review or meta-analysis. Syntax, search fields, and subject headings were adapted for each database. No date or language restrictions were applied at the search stage.

Terms describing digital interventions were informed by previous systematic reviews and meta-reviews^{12, 14} and covered internet- and web-based interventions, computerized interventions, mobile applications, e-health and m-health, text messaging, conversational agents, video consultations, virtual reality, and therapeutic games. Before the formal searches were run, we checked whether the strategy retrieved several prespecified reviews closely related to the research question^{19–21, 58}. When a benchmark review was not retrieved, its title, abstract, keywords, and database subject headings were examined to refine the search terms without changing the eligibility criteria.

We additionally screened the reference lists of included reviews and relevant umbrella reviews and conducted forward citation tracking. Records were deduplicated before formal screening. Grey literature was not searched separately, and exploratory feasibility searches were not included in the formal study-selection counts. Full database-specific strategies, platforms, search dates, and yields are provided in Supplementary Table S1.

Study selection and data extraction

After deduplication, Y.Q. and Q.S. divided the records equally and screened their assigned titles and abstracts. To assess consistency in applying the eligibility criteria, each reviewer randomly cross-checked 20% of the other reviewer's decisions (685 records each; 1,370 records in total). Reports selected for full-text assessment were independently evaluated by both reviewers against the prespecified eligibility criteria. For each excluded full-text report, one primary reason was recorded according to the first decisive criterion not met. Disagreements were resolved through discussion and, when necessary, adjudication by J.H. Initial agreement at the full-text stage was quantified using raw percentage agreement and Cohen's κ . Study selection was reported according to PRISMA 2020, and excluded full-text reports with their primary reasons are listed in Supplementary Table S3⁵⁵.

Data were collected using a standardized form developed in advance and refined through pilot extraction. Y.Q. extracted the data, and Q.S. checked every item against the review articles, supplementary materials, protocols, and registration records. Extracted information included review and search characteristics; populations; interventions and comparators; outcomes and assessment times; studies included in each meta-analysis; sample sizes; effect estimates and measures of uncertainty; statistical methods and heterogeneity; and reported assessments of primary-

study risk of bias and evidence certainty. The extraction fields were informed by Cochrane Handbook guidance for overviews of reviews⁵¹.

When reports contained conflicting information, we prioritized values corresponding to the target analysis that were clearly defined and traceable to a specific source. Primary RCT reports were consulted when necessary to verify participant ages, comparator content, meta-analysis membership, randomized sample sizes, or whether multiple publications originated from the same trial. Information that could not be verified reliably was recorded as not reported rather than inferred. Disagreements concerning eligibility, effect selection, comparator classification, or trial linkage were resolved through discussion and, when necessary, adjudication by J.H.

Outcome definition and effect selection

The strict primary estimand was the between-group postintervention difference in depressive symptom severity between a digital psychological intervention and a passive or nontherapeutic control among people aged 10–25 years. Continuous effects were harmonized as standardized mean differences (SMDs), with negative values favoring the digital intervention. Effect estimates and standard errors were extracted from the source reviews or derived from reported confidence intervals or other convertible statistics in accordance with Cochrane Handbook guidance⁵⁹.

When a review reported multiple candidate effects, we first applied the prespecified population, intervention, comparator, outcome, assessment time, and study-design criteria. If more than one effect remained eligible, we prioritized, in sequence, the effect with the most clearly defined target-age population, the most direct measure of depressive symptoms, an extractable between-group postintervention estimate, and verifiable analysis membership. Supplementary Table S2 documents effect selection, source locations, direction harmonization, and eligibility decisions.

Standardized between-group change-score effects were reserved for sensitivity analysis and were not included in the strict primary model. Effects against active treatments were analyzed separately from effects against passive or nontherapeutic controls. Mixed-age effects without separately extractable results for people aged 10–25 years entered the extended-evidence analysis only if they met the prespecified 75% rule.

Each review contributed no more than one effect to the strict primary model. We did not treat correlated effects from the same review as independent observations or combine mutually exclusive effects to create an overall estimate not reported by the source review. Candidate effects were retained in the extraction ledger to document selection, harmonization, and exclusion decisions.

Methodological quality of included reviews

We assessed the methodological quality of included systematic reviews using AMSTAR 2⁵². Y.Q. and Q.S. independently completed initial judgments for all 16 items using the official checklist and guidance and consulted review articles, supplementary materials, publicly available protocols, and registration records. Methods that were not reported and could not be verified from related materials were not assumed to have been performed. Items were rated as “Yes,” “Partial Yes,” or “No,” with “Not Applicable” used where appropriate.

Raw percentage agreement and unweighted Cohen’s κ were calculated from the initial independent item-level judgments before consensus. Disagreements were resolved by rechecking the source materials and official decision rules, with J.H. adjudicating unresolved cases. Final AMSTAR 2 ratings reflected consensus judgments.

Overall confidence was determined from weaknesses in critical domains rather than a numerical total score and was classified as high, moderate, low, or critically low according to AMSTAR 2 guidance. Review-level methodological confidence informed interpretation but was not used as an automatic exclusion criterion for the strict primary analysis. We planned a quality-restricted analysis excluding critically low-quality reviews; if too few effects remained for reliable pooling, the remaining evidence was to be described without quantitative synthesis.

Identification of primary studies and assessment of overlap

To identify primary evidence shared across reviews, we extracted eligible randomized trials from review articles, supplementary materials, forest plots, and study-characteristics tables. Trial reports, registrations, and companion publications were consulted when necessary to verify trial identity, participant characteristics, interventions, comparators, and analytic samples.

Primary-outcome reports, follow-up reports, subgroup analyses, and process evaluations arising from the same randomized trial were linked as a report family and counted as one independent RCT cluster. Linkage considered trial names, author groups, registration numbers, study sites, recruitment periods, sample characteristics, numbers randomized, and intervention arms. Matching or differing author–year labels alone were not regarded as sufficient evidence that reports originated from the same or different trials. Reports that could not be linked reliably were retained as separate candidate records, with the uncertainty documented in the audit ledger.

Overlap was evaluated using a review-by-RCT citation matrix. The corrected covered area (CCA) was calculated separately for each quantitative model as:

$$CCA = \frac{N-r}{rc-r} \times 100\%$$

where N is the total number of review–RCT occurrences in the matrix, r is the number of independent RCT clusters, and c is the number of systematic reviews in the analysis³⁴. Multiple effects from the same trial within a review contributed only one RCT occurrence. CCA was calculated from the actual reviews and RCT members entering each model rather than substituting the overall overlap across all included reviews. Values of 0%–5%, > 5%–10%, > 10%–15%, and > 15% were interpreted as slight, moderate, high, and very high overlap, respectively³⁴.

Statistical analysis

The unit of analysis in the meta-meta-analysis was the pooled effect reported by a systematic review. Depression outcomes were harmonized so that negative values favored the digital psychological intervention. When a source review used the opposite direction, the effect estimate and its confidence limits were reversed together. Standard errors were extracted directly when available or derived from two-sided 95% confidence intervals as:

$$SE = \frac{CI_{upper} - CI_{lower}}{2z_{0.975}} = \frac{CI_{upper} - CI_{lower}}{3.92}$$

Given clinical and methodological differences in populations, intervention formats, human support, comparators, and outcome scales, all quantitative syntheses used inverse-variance-weighted random-effects models. Between-review variance τ^2 was estimated using restricted maximum likelihood, and pooled standard errors, confidence intervals,

and significance tests were adjusted using the Knapp–Hartung method. This approach more fully reflects uncertainty when the number of review-level effects is limited.

For each model, we report the pooled effect and 95% confidence interval, τ^2 , Cochran's Q with its degrees of freedom and p value, and I^2 . A 95% prediction interval was also calculated to indicate the range in which a true effect might occur in a future evidence setting comparable to the included reviews. Prediction intervals were interpreted alongside pooled effects and between-review heterogeneity, rather than judging consistency solely from I^2 or the Q test. We prespecified that prediction intervals would be reported only for models containing at least five review-level effects; because intervals may be unstable with few effects, their interpretation considered effect count and heterogeneity⁵⁹.

All statistical tests were two-sided, with $\alpha = 0.05$. Effect magnitude was interpreted from the point estimate, confidence interval, prediction interval, and clinical and methodological context rather than statistical significance alone. Analyses were conducted in R version 4.3.3 using the metafor package⁶⁰.

Primary, supporting, and sensitivity analyses

The strict primary analysis included review-level effects meeting all of the following criteria: participants aged 10–25 years; a digital psychological intervention; a passive or nontherapeutic comparator; depressive symptoms measured at postintervention; a randomized between-group postintervention effect; and sufficiently verifiable analysis membership. Each review contributed no more than one effect to the strict primary model.

Sensitivity to effect definition was examined by adding standardized between-group change-score effects that otherwise met the strict primary criteria. We also conducted a leave-one-review-out analysis, sequentially omitting each review and re-estimating the pooled effect, heterogeneity, and prediction interval to determine whether the result was driven by any single review.

Effects comparing digital interventions with controls containing substantive therapeutic components were examined in a separate active-control analysis and were not pooled with passive-control effects. Mixed-age effects that met the prespecified 75% threshold but lacked a separately extractable result for participants aged 10–25 years were evaluated in an extended-evidence sensitivity analysis.

Differences according to intervention type, degree of human support, prevention or treatment purpose, and follow-up time were considered for quantitative subgroup analysis only when sufficiently numerous, consistently defined, and independent review-level effects were available. Otherwise, the evidence was summarized narratively or displayed in structured evidence plots rather than pooling clinically or methodologically noncomparable estimates.

Methodological quality was not used as an automatic exclusion criterion for the strict primary analysis. We prespecified a quality-restricted analysis excluding reviews rated critically low by AMSTAR 2. If too few reviews remained for a reliable random-effects synthesis, the remaining estimates were described without quantitative pooling.

Small-study effects and publication bias

Publication bias was evaluated for the strict primary analysis. Funnel-plot asymmetry and regression-based tests are generally considered informative only when a meta-analysis contains at least 10 sufficiently independent effects⁵⁹. The strict primary analysis contained nine review-level effects, and several reviews shared underlying RCTs, further

weakening the independence required for formal asymmetry testing. We therefore did not perform Egger regression or another formal funnel-plot asymmetry test.

Instead, publication bias was assessed narratively using funnel plots, regression tests, trim-and-fill analyses, searches for unpublished evidence, and related judgments reported by the source reviews. These findings were considered together with the coverage of the present search and the feasibility of formal testing in the GRADE assessment. The inability to perform a formal asymmetry test was treated as a limitation of the evidence rather than evidence that publication bias was absent.

Certainty of evidence

We used GRADE to assess certainty in the strict primary outcome: the postintervention effect of digital psychological intervention versus passive or nontherapeutic control on depressive symptoms⁶¹. Although the quantitative synthesis used review-level pooled effects, risk-of-bias information was traced to the primary RCTs underlying the strict primary model. Because the evidence originated from randomized trials, certainty initially started at high and was considered for downgrading across risk of bias, inconsistency, indirectness, imprecision, and publication bias.

Risk of bias was assessed from information on randomization, deviations from intended interventions, missing outcome data, outcome measurement, and selective reporting in the primary RCTs; AMSTAR 2 ratings of source reviews were not substituted for primary-study risk of bias. Inconsistency was assessed from effect directions, the range of estimates, I^2 , Cochran's Q , and the prediction interval. Indirectness was evaluated against the prespecified population, intervention, comparator, outcome, and time point. Imprecision was considered in light of the 95% confidence interval, prediction interval, estimate stability, and the size of the underlying evidence base. Publication bias was assessed from the feasibility of formal testing, relevant analyses in source reviews, search coverage, and discussion of unpublished evidence.

Y.Q. and Q.S. independently completed the GRADE assessment. Disagreements were resolved by reviewing the supporting evidence and through discussion, with J.H. adjudicating when required. Overall certainty was classified as high, moderate, low, or very low. Domain judgments and reasons for downgrading are reported in the Summary of Findings table and Supplementary Table S10.

Declarations

Author Contribution

Q.Y. and Z.R. conceptualized the study. Q.Y. developed the methodology, conducted the investigation, curated the data, performed the formal analyses, prepared the visualizations, and drafted the manuscript. Q.S. contributed to the methodology, investigation, data curation, and validation. J.H. contributed to the methodology, validation, adjudication, and supervision. P.Y. and S.C. contributed to the investigation and data collection and curation. Z.R. supervised and administered the project and reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgement

We thank the authors of the systematic reviews and primary studies included in this work for making their research and supplementary materials available.

Data Availability

The review-level data underlying the analyses, effect-selection and eligibility records, AMSTAR 2 item-level ratings, GRADE judgments, and primary-study citation inventory are provided in the Supplementary Information. Full-text articles obtained through institutional subscriptions are third-party copyrighted materials and are not redistributed by the authors; they may be accessed through the respective publishers subject to their access conditions. Additional intermediate screening, extraction, and audit files containing reviewer working annotations are available from the corresponding author upon reasonable request.

Data availability

The review-level data underlying the analyses, effect-selection and eligibility records, AMSTAR 2 item-level ratings, GRADE judgments, and primary-study citation inventory are provided in the Supplementary Information. Full-text articles obtained through institutional subscriptions are third-party copyrighted materials and are not redistributed by the authors; they may be accessed through the respective publishers subject to their access conditions. Additional intermediate screening, extraction, and audit files containing reviewer working annotations are available from the corresponding author upon reasonable request.

Code availability

The analysis code used in this study can be obtained from the corresponding author upon reasonable request.

Human Ethics and Consent to Participate declarations

not applicable.

References

1. Lu, B., Li, L., Lin, L. & Su, X. Global burden of depression or depressive symptoms in children and adolescents: A systematic review and meta-analysis. *J. Affect. Disord.* 354, 553–562 (2024).
2. Thapar, A., Eyre, O., Patel, V. & Brent, D. Depression in young people. *Lancet* 400, 617–631 (2022).
3. Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M. & Merikangas, K. R. Major depression in the National Comorbidity Survey–Adolescent Supplement: Prevalence, correlates, and treatment. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 37–44 (2015).
4. Clayborne, Z. M., Varin, M. & Colman, I. Systematic review and meta-analysis: Adolescent depression and long-term psychosocial outcomes. *J. Am. Acad. Child Adolesc. Psychiatry* 58, 72–79 (2019).
5. Copeland, W. E., Alaie, I., Jonsson, U. & Shanahan, L. Associations of childhood and adolescent depression with adult psychiatric and functional outcomes. *J. Am. Acad. Child Adolesc. Psychiatry* 60, 604–611 (2021).
6. Fergusson, D. M., Horwood, L. J., Ridder, E. M. & Beautrais, A. L. Subthreshold depression in adolescence and mental health outcomes in adulthood. *Arch. Gen. Psychiatry* 62, 66–72 (2005).
7. Klein, D. N., Shankman, S. A., Lewinsohn, P. M. & Seeley, J. R. Subthreshold depressive disorder in adolescents: Predictors of escalation to full-syndrome depressive disorders. *J. Am. Acad. Child Adolesc. Psychiatry* 48, 703–710 (2009).
8. Ssegonga, R. et al. Indicated preventive interventions for depression in children and adolescents: A meta-analysis and meta-regression. *Prev. Med.* 118, 7–15 (2019).

9. American Psychological Association. Clinical practice guideline for the treatment of depression across three age cohorts. <https://www.apa.org/depression-guideline> (2019).
10. National Institute for Health and Care Excellence. Depression in children and young people: Identification and management (NICE Guideline NG134). <https://www.nice.org.uk/guidance/ng134> (2019).
11. Radez, J. et al. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. *Eur. Child Adolesc. Psychiatry* 30, 183–211 (2021).
12. Hollis, C. et al. Annual research review: Digital health interventions for children and young people with mental health problems—A systematic and meta-review. *J. Child Psychol. Psychiatry* 58, 474–503 (2017).
13. Liverpool, S. et al. Updates on digital mental health interventions for children and young people: Systematic overview of reviews. *Eur. Child Adolesc. Psychiatry* 34, 2961–2974 (2025).
14. Zhang, M. et al. Assessing the effectiveness of internet-based interventions for mental health outcomes: An umbrella review. *Gen. Psychiatry* 37, e101355 (2024).
15. Lehtimäki, S., Martić, J., Wahl, B., Foster, K. T. & Schwalbe, N. Evidence on digital mental health interventions for adolescents and young people: Systematic overview. *JMIR Ment. Health* 8, e25847 (2021).
16. Pennant, M. E. et al. Computerised therapies for anxiety and depression in children and young people: A systematic review and meta-analysis. *Behav. Res. Ther.* 67, 1–18 (2015).
17. Ebert, D. D. et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. *PLoS ONE* 10, e0119895 (2015).
18. Christ, C. et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: Systematic review and meta-analysis. *J. Med. Internet Res.* 22, e17831 (2020).
19. Eilert, N., Wogan, R., Leen, A. & Richards, D. Internet-delivered interventions for depression and anxiety symptoms in children and young people: Systematic review and meta-analysis. *JMIR Pediatr. Parent.* 5, e33551 (2022).
20. Välimäki, M., Anttila, K., Anttila, M. & Lahti, M. Web-based interventions supporting adolescents and young people with depressive symptoms: Systematic review and meta-analysis. *JMIR Mhealth Uhealth* 5, e180 (2017).
21. Wickersham, A., Barack, T., Cross, L. & Downs, J. Computerized cognitive behavioral therapy for treatment of depression and anxiety in adolescents: Systematic review and meta-analysis. *J. Med. Internet Res.* 24, e29842 (2022).
22. Harrer, M. et al. Internet interventions for mental health in university students: A systematic review and meta-analysis. *Int. J. Methods Psychiatr. Res.* 28, e1759 (2019).
23. Wang, Q., Zhang, W. & An, S. A systematic review and meta-analysis of internet-based self-help interventions for mental health among adolescents and college students. *Internet Interv.* 34, 100690 (2023).
24. Thabrew, H. et al. E-health interventions for anxiety and depression in children and adolescents with long-term physical conditions. *Cochrane Database Syst. Rev.* 2018, CD012489 (2018).
25. Chen, T. H., Chu, G., Pan, R.-H. & Ma, W.-F. Effectiveness of mental health chatbots in depression and anxiety for adolescents and young adults: A meta-analysis of randomized controlled trials. *Expert Rev. Med. Devices* 22, 233–241 (2025).
26. Feng, Y. et al. Effectiveness of AI-driven conversational agents in improving mental health among young people: Systematic review and meta-analysis. *J. Med. Internet Res.* 27, e69639 (2025).
27. Leech, T., Dorstyn, D., Taylor, A. & Li, W. Mental health apps for adolescents and young adults: A systematic review of randomised controlled trials. *Child. Youth Serv. Rev.* 127, 106073 (2021).

28. Magwood, O. et al. Mobile apps to reduce depressive symptoms and alcohol use in youth: A systematic review and meta-analysis. *Campbell Syst. Rev.* 20, e1398 (2024).
29. Reangsing, C., Abdullahi, S. G. & Schneider, J. K. Effects of online mindfulness-based interventions on depressive symptoms in college and university students: A systematic review and meta-analysis. *J. Integr. Complement. Med.* 29, 292–302 (2023).
30. Townsend, C. et al. The effectiveness of gaming interventions for depression and anxiety in young people: Systematic review and meta-analysis. *BJPsych Open* 8, e25 (2022).
31. Chen, B. et al. Comparative effectiveness and acceptability of internet-based psychological interventions on depression in young people: A systematic review and network meta-analysis. *BMC Psychiatry* 25, 321 (2025).
32. Li, J. et al. Optimal approaches for preventing depressive symptoms in children and adolescents based on the psychosocial interventions: A Bayesian network meta-analysis. *J. Affect. Disord.* 280, 364–372 (2021).
33. Wang, H. et al. Comparative efficacy of digital health interventions for depression and anxiety symptoms in adolescents and young adults: A systematic review and Bayesian network meta-analysis. *Child Adolesc. Psychiatry Ment. Health* 20, 25 (2026).
34. Pieper, D., Antoine, S.-L., Mathes, T., Neugebauer, E. A. M. & Eikermann, M. Systematic review finds overlapping reviews were not mentioned in every other overview. *J. Clin. Epidemiol.* 67, 368–375 (2014).
35. Madrid-Cagigal, A. et al. Digital mental health interventions for university students with mental health difficulties: A systematic review and meta-analysis. *Early Interv. Psychiatry* 19, e70017 (2025).
36. Garrido, S. et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. *Front. Psychiatry* 10, 759 (2019).
37. Hatoum, A. H. et al. Digital single-session interventions for depression and self-injurious thoughts and behaviours in young people: A systematic review and meta-analysis. *eClinicalMedicine* 90, 103601 (2025).
38. Lin, J. T., Syu, J. Y. & Lin, E. C. L. Effectiveness of mindfulness-based apps in mitigating psychological distress among youth: A systematic review and meta-analysis. *Mindfulness* (2026). <https://doi.org/10.1007/s12671-026-02898-4>
39. Liu, P. et al. The acceptability and effects of internet-based cognitive behavioral therapy on depressive symptoms and remission in 13- to 17-year-old adolescents: A systematic review and meta-analysis. *J. Evid. Based Med.* 17, 740–757 (2024).
40. Noh, D., Park, H. & Shim, M.-S. Website and mobile application-based interventions for adolescents and young adults with depression: A systematic review and meta-analysis. *J. Korean Acad. Psychiatr. Ment. Health Nurs.* 32, 78–91 (2023).
41. Noh, D. & Kim, H. Effectiveness of online interventions for the universal and selective prevention of mental health problems among adolescents: A systematic review and meta-analysis. *Prev. Sci.* 24, 353–364 (2023).
42. Sathe, A., Nainani, A., Elaraby, A., Jones, A. L. & Lowe, M. Effectiveness of mobile-based cognitive behavioural therapy for adolescent depression: A systematic review and meta-analysis. *Aust. N. Z. J. Psychiatry* 60, 633–642 (2026).
43. Tian, E. et al. Effectiveness of telepsychiatry interventions for youth with depressive and/or anxiety disorders: A systematic review with meta-analysis. *Clin. Child Psychol. Psychiatry* 30, 537–561 (2025).
44. Wu, Y. et al. Efficacy of internet-based cognitive-behavioral therapy for depression in adolescents: A systematic review and meta-analysis. *Internet Interv.* 34, 100673 (2023).
45. Bevilacqua, L. et al. Effectiveness of technology-assisted vs face-to-face cognitive behavioural therapy for anxiety and depression in children and young people: A systematic review and meta-analysis. *Clin. Child Psychol.*

Psychiatry 29, 1349–1364 (2024).

46. Bryant, B. R., Sisk, M. R. & McGuire, J. F. Efficacy of gamified digital mental health interventions for pediatric mental health conditions: A systematic review and meta-analysis. *JAMA Pediatr.* 178, 1136–1146 (2024).
47. Grist, R., Stallard, P., Croker, A. & Denne, M. Technology delivered interventions for depression and anxiety in children and adolescents: A systematic review and meta-analysis. *Clin. Child Fam. Psychol. Rev.* 22, 147–171 (2019).
48. Takizawa, Y. et al. Are universal digital mental health interventions effective for promoting mental health outcomes among children and adolescents? A meta-analysis of effects and moderators. *J. Telemed. Telecare* 32, 391–411 (2026).
49. Ye, X. et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: A systematic review and meta-analysis. *BMC Health Serv. Res.* 14, 313 (2014).
50. Dietvorst, E. et al. Effective elements of eHealth interventions for mental health and well-being in children and adolescents: A systematic review. *Digit. Health* 10 (2024). <https://doi.org/10.1177/20552076241294105>
51. Pollock, M., Fernandes, R. M., Becker, L. A., Pieper, D. & Hartling, L. Chapter V: Overviews of reviews. In *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5 (eds Higgins, J. P. T. et al.) (Cochrane, 2024).
52. Shea, B. J. et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358, j4008 (2017).
53. Guyatt, G. H. et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 64, 383–394 (2011).
54. Gates, M. et al. Reporting guideline for overviews of reviews of healthcare interventions: Development of the PRIOR statement. *BMJ* 378, e070849 (2022).
55. Page, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372, n71 (2021).
56. Storebø, O. J. et al. Methylphenidate for children and adolescents with attention deficit hyperactivity disorder (ADHD). *Cochrane Database Syst. Rev.* 2015, CD009885 (2015).
57. Firth, J. et al. The efficacy and safety of nutrient supplements in the treatment of mental disorders: A meta-review of meta-analyses of randomized controlled trials. *World Psychiatry* 18, 308–324 (2019).
58. Dülsen, P. & Baumeister, H. Internet- and mobile-based anxiety and depression interventions for children and adolescents: Efficacy and negative effects—A systematic review and meta-analysis. *Eur. Child Adolesc. Psychiatry* 34, 101–121 (2025).
59. Higgins, J. P. T. et al. (eds) *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5 (Cochrane, 2024).
60. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48 (2010).
61. Guyatt, G. H. et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336, 924–926 (2008).

Figures

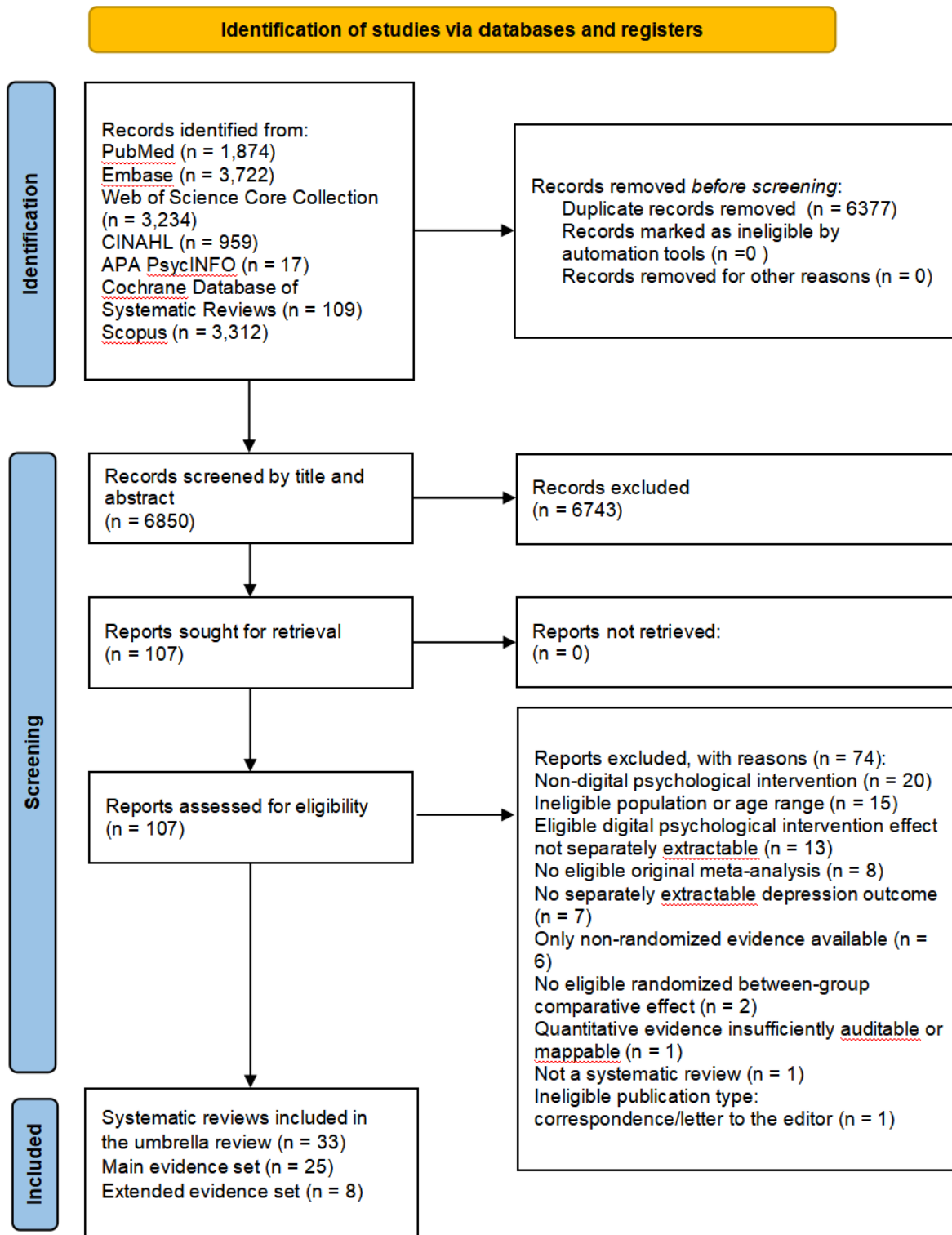
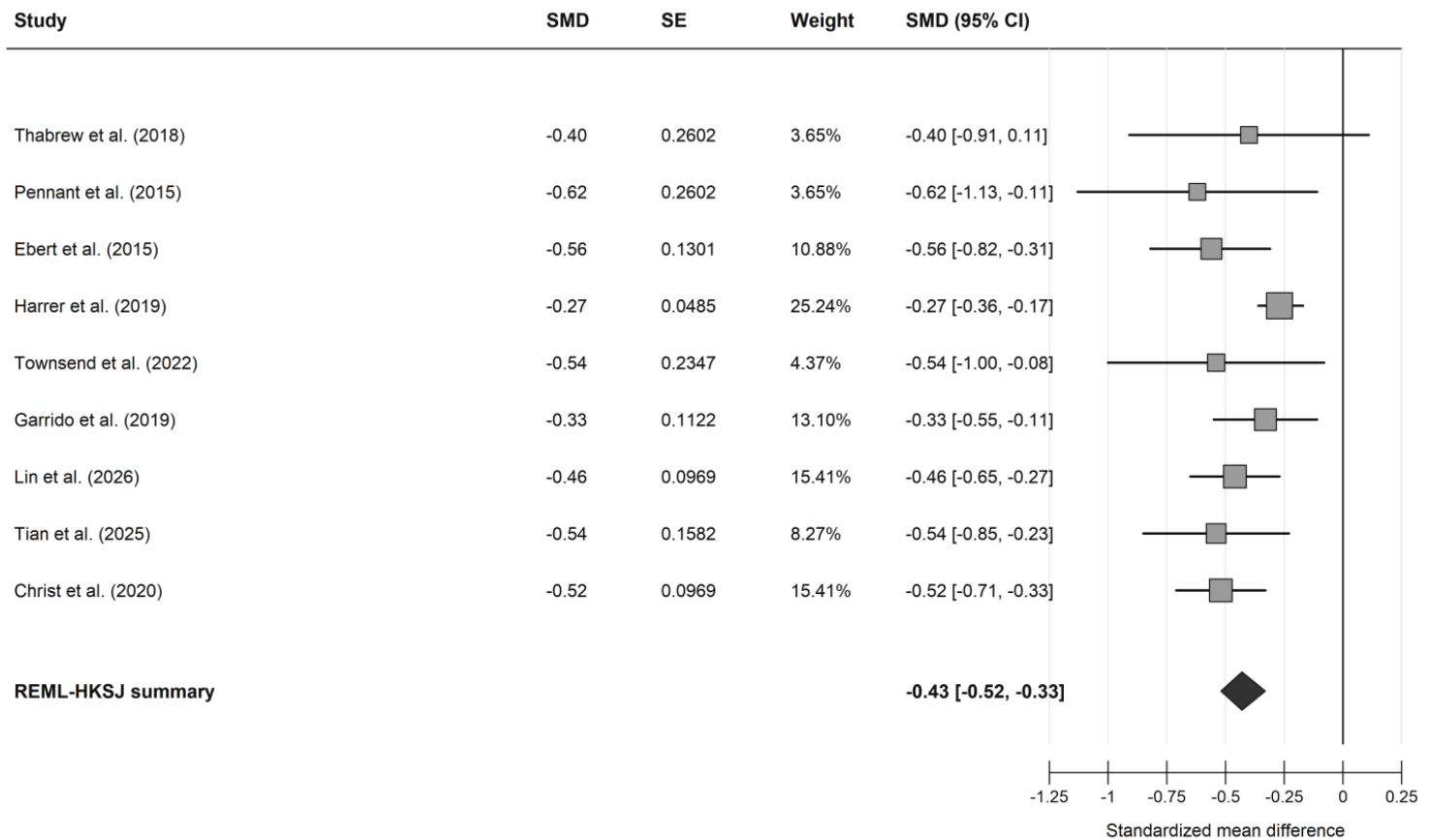


Figure 1

PRISMA 2020 flow diagram of study selection.



Heterogeneity: $Q = 12.36$, $df = 8$, $p = .136$; $I^2 = 35.27\%$; $\tau^2 = 0.0087$; 95% PI -0.67 to -0.19; CCA = 4.23%.

Figure 2

Strict primary meta-meta-analysis of postintervention depressive symptoms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryResultsandAppendices.docx](#)