

Digital Cognitive Bias Modification for Interpretation: Multilevel Meta-Analysis of Interpretation Bias, Anxiety, and Depression Outcomes

Kaan Alp Karamanli

King's College London

Jonas Everaert

Tilburg University

Xiadong Wu

King's College London

Mina Şahintürk

King's College London

Zeqi Jia

King's College London

Katie Lam

King's College London

Daniel Stahl

King's College London

Sukhi Shergill

King's College London

Jenny Yiend

`jenny.yiend@kcl.ac.uk`

King's College London

Article

Keywords:

Posted Date: October 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11071288/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

Cognitive bias modification for interpretation (CBM-I) is scalable through digital delivery, but its effects on anxiety and depression remain uncertain. We systematically reviewed digital CBM-I studies reporting interpretation bias, anxiety, or depression and synthesised dependent effect sizes using three-level random-effects models. 73 trials with 10,870 participants were included. At post-treatment, CBM-I significantly improved interpretation bias ($g = -0.54$, 95% CI -0.64 to -0.43), anxiety ($g = -0.19$, 95% CI -0.27 to -0.12), and depression ($g = -0.16$, 95% CI -0.25 to -0.07). Effects remained significant at follow-up (ranging between 2–52 weeks) with similar magnitude. Clinical status and stimulus modality moderated post-treatment depression, with larger effects in clinical samples and with auditory delivery. Digital CBM-I produces moderate cognitive and modest but significant symptom benefits, supporting further evaluation in clinical populations, but the predominance of high-risk studies, early-stage trials and limited follow-up evidence does not support readiness for routine implementation.

Introduction

Despite decades of research and numerous intervention efforts, depression and anxiety remain leading causes of global disability.¹ Gold-standard treatments prescribed for depression and anxiety (e.g., Cognitive-Behavioural Therapy) remain inaccessible for many individuals mostly due to limited number of trained clinicians.² Digital interventions that can be self-administered and delivered at scale may help address this persistent access gap. Among these, digital cognitive bias modification for interpretation (CBM-I), designed to modify negative interpretation bias (i.e., tendency to interpret ambiguity in a negative way), has been proposed as a promising, scalable intervention,³ as it is evidence-based and grounded in cognitive theory.^{4,5} Indeed, compared to CBMs targeting other cognitive biases (e.g., attention, approach-avoidance), CBM-I showed the most consistent effects both for depression and anxiety⁶ and has been proposed for implementation in routine mental healthcare.³

However, several issues prevent its large-scale implementation. To date, CBM-I was either self-administered through web-based platforms or mobile apps (defined here as 'digital CBM-I'), or administered in-person involving researchers, generally in laboratory settings.⁷ Among these two administration methods, digital CBM-I holds most promise for scalability, yet the meta-analytic studies to date have pooled digital and laboratory-administered CBM-I studies together⁶, despite some evidence showing that digital versions might be less potent⁸. Since cost-effectiveness claims as well as scalability of CBM-I may only be valid when delivered digitally, decisive evidence on the efficacy of digital CBM-I is needed. Second, it remains unclear whether digital CBM-I is less effective for clinical populations, as some researchers have promoted its preventive use in healthy populations.⁷ However, recent studies demonstrated significant effects of CBM-I even when delivered digitally⁹ and in clinical groups,¹⁰ warranting an updated meta-analytic synthesis. Moreover, other important candidate moderators, such as age, training paradigm (i.e., ambiguous scenarios vs other), stimulus modality (i.e., text versus auditory), and imagery, also await examining to uncover what works for whom in digital format, as meta-

analyses to date have investigated these by pooling laboratory-delivery and digital-delivery studies together.^{6–8,11} Lastly, previous meta-analyses mostly lacked follow-up outcomes^{6,8,12}, despite that the preservation of any symptom and cognitive improvements is of crucial importance for real-life implementation.

In addition to lacking evidence specific to digital versions, meta-analyses to date show some other notable limitations. First, most used two-level meta-analytic methods,^{6–8,11} assuming independence of effect sizes, hence forcing researchers to choose some conditions and measures over others, or aggregate them. This could introduce bias and reduce validity, as many studies include multiple CBM-I/control conditions and measures per outcome (e.g.,^{9,13}). Moreover, bias in these meta-analyses has been further compounded by the exclusion of grey literature.^{6–8,11} Second, multiple meta-analyses have cast doubt on CBM-I's clinical efficacy for depression, compared to anxiety,^{6,11,14,15} in contrast with many recent studies yielding significant effects for depression as well as for anxiety, contrasting with these previous meta-analytic findings.¹⁶ This may be due to recent efforts to tailor CBM-I for depression,^{17–19} but these might not have been captured in earlier meta-analyses.

In this systematic review and multi-level meta-analysis, we synthesised evidence from 74 digital CBM-I studies to quantify its effects on interpretation bias, anxiety, and depression at post-treatment and follow-up. We examined whether effects are maintained over time, whether they differed across the levels of candidate moderators including but not limited to clinical status, comparator type, stimulus modality, training paradigm, and imagery. By employing a three-level framework for between-group comparisons, this study avoided arbitrary prioritisation or aggregation of outcomes and conditions²⁰ and provided the most comprehensive assessment to date of whether digital CBM-I produces significant and durable changes in cognitive bias and emotional symptoms.

Results

Study Selection and Characteristics

A total of 674 records were identified through databases (n = 608), trial registers (n = 21), and dissertations (n = 45). After automatic duplicate removal (n = 325) and exclusion of studies included in earlier searches (n = 5), 344 records were screened by title and abstract, of which 226 were excluded. The remaining 118 full-text reports were assessed for eligibility, and 47 were excluded for reasons detailed in Fig. 1 (see Table S9 for exclusion reasons for each report). We contacted 2 authors for completely missing data and received no responses. 21 other authors were contacted for additional data (e.g., age means of groups, gender percentage; data not essential for the primary meta-analyses) and 14 responded. Inter-rater reliability calculated using Cohen's Kappa coefficient was 0.80 for the abstract screening and 0.71 for the full-text screening, both indicating substantial agreement.

74 studies from 71 reports were eligible for the systematic review. The difference in numbers between studies and reports is because some reports contained multiple independent experiments. One study

(one report) was excluded from the meta-analyses due to missing essential data, leaving 73 studies contributing at least to one of the meta-analytic syntheses ($N_{T0/BASELINE} = 10,870$; $N_{T1/POST-TREATMENT} = 8,310$). Study characteristics for each included study are given in the supplement (Table S5). Specifically, 65 were published and 8 unpublished. Most were recent, with 59 dated after 2015 and more than half ($n = 41$) dated after 2020. Samples were mainly adult ($n = 62$), with 10 in minors and one mixed-age. Clinical status was clinical in 21 studies, subclinical in 43, and healthy in nine. Anxiety disorder diagnoses ($n = 43$) were followed by other disorders ($n = 14$), depression ($n = 5$), and comorbid anxiety and depression ($n = 3$), with no diagnosis in nine studies (Hirsch et al., 2018²¹ included two diagnostic groups, generalised anxiety disorder and depression). Delivery was mostly web-based ($n = 61$), with 12 mobile-app trials. Control conditions most often involved sham (i.e., 50% positive/50% negative; $n = 17$) or neutral CBM-I ($n = 17$), with smaller numbers using waiting list/treatment as usual ($n = 11$), unresolved ambiguity ($n = 8$), no-task ($n = 7$), other/unspecified ($n = 14$), and negative CBM-I ($n = 2$), with 10 studies having no control group. Follow-up data from 41 studies were included in at least one quantitative synthesis ($N_{T2/FOLLOW-UP} = 3,637$). Time from baseline to the follow-up assessment ranged from 2–29 weeks for interpretation bias ($M = 8.45$, $SD = 6.44$), and 2–52 weeks for both anxiety ($M = 11.60$, $SD = 9.78$) and depression ($M = 11.80$, $SD = 9.96$). At post-treatment, the meta-analyses included 57 studies contributing 324 effect sizes for negative interpretation bias, 49 studies contributing 212 effect sizes for anxiety, and 37 studies contributing 69 effect sizes for depression. At follow-up, the corresponding numbers were 16 studies contributing 65 effect sizes for negative interpretation bias, 26 studies contributing 89 effect sizes for anxiety, and 25 studies contributing 49 effect sizes for depression. Therefore, post-treatment analyses were mostly based on a sufficiently large number of studies to permit stable three-level variance estimation (≈ 40 – 50).²² However, follow-up estimates were retained to provide preliminary insight into longer-term effects.

Figure 1. PRISMA flow diagram describing the identification, screening, and inclusion of the relevant reports in the systematic review.

Risk of Bias

Risk of bias was generally elevated, with only six studies rated low risk overall. Among studies contributing post-treatment effects, high-risk ratings predominated for interpretation bias (37/57, 64.9%), anxiety (33/49, 67.3%), and depression (23/37, 62.2%); low-risk studies accounted for only 5/57 (8.8%), 5/49 (10.2%), and 4/37 (10.8%), respectively.

Publication Bias

Visual inspection of the funnel plots produced mixed indications, while modified Egger tests indicated funnel-plot asymmetry only for follow-up anxiety ($b = 0.88$, $SE = 0.08$, $p < .001$; Figure S1 and Table S8). Under the tested significance-selection model, finite S-values could not be estimated for any outcome other than for post-treatment interpretation bias, because no finite selection ratio nullified the pooled

effects. For post-treatment interpretation bias, the value was 2.48, meaning that affirmative studies would need to be 2.48 times more likely to be published for the confidence intervals of the effect to include zero. Leave-one-cluster-out analyses indicated that no individual study materially influenced the pooled results (Figures S3–S8).

Primary Meta-Analyses: Effects of Digital CBM-I on Interpretation Bias, Anxiety, and Depression at Post-Treatment and Follow-Up

See Table 1 for the summary of results, including the number of studies and effect sizes and heterogeneity per outcome. Overall, digital CBM-I was associated with significantly greater improvements than control conditions in interpretation bias, anxiety, and depression at both post-treatment and follow-up.

For interpretation bias, significant effects were found at post-treatment ($g = -0.54$, 95% CI $[-0.64, -0.43]$) and follow-up ($g = -0.46$, 95% CI $[-0.58, -0.34]$; both $p < .0001$). Estimated heterogeneity was predominantly between studies at both post-treatment (between-study $\sigma^2 = 0.102$; within-study $\sigma^2 < 0.001$) and follow-up (between-study $\sigma^2 = 0.015$; within-study $\sigma^2 < 0.001$).

Anxiety effects were significant at post-treatment ($g = -0.19$, 95% CI $[-0.27, -0.12]$) and follow-up ($g = -0.23$, 95% CI $[-0.31, -0.15]$; both $p < .0001$). Estimated heterogeneity was again predominantly between studies at post-treatment (between-study $\sigma^2 = 0.024$; within-study $\sigma^2 < 0.001$) and follow-up (between-study $\sigma^2 = 0.005$; within-study $\sigma^2 < 0.001$).

Depression effects were significant at post-treatment ($g = -0.16$, 95% CI $[-0.25, -0.07]$, $p = .0012$) and follow-up ($g = -0.17$, 95% CI $[-0.26, -0.08]$, $p = .0012$). Estimated heterogeneity was predominantly between studies at post-treatment (between-study $\sigma^2 = 0.026$; within-study $\sigma^2 < 0.001$) and follow-up (between-study $\sigma^2 = 0.006$; within-study $\sigma^2 < 0.001$).

After Benjamini–Hochberg correction across the six primary pooled-effect tests, all effects remained statistically significant (BH-adjusted $p < .0001$ for interpretation bias and anxiety at both time points and BH-adjusted $p = .0012$ for depression at both time points). Supplementary within-group analyses of CBM-I arms at post-treatment were consistent with the primary between-group findings, showing significant baseline-to-post-treatment improvements in interpretation bias ($g = -0.83$, 95% CI $[-0.95, -0.72]$), anxiety ($g = -0.54$, 95% CI $[-0.63, -0.44]$), and depression ($g = -0.45$, 95% CI $[-0.56, -0.35]$; all $p < .0001$; Table S4).

All pooled effects remained statistically significant after outlier exclusion (Supplementary Section S3). Estimated between-study heterogeneity was lower at follow-up than at post-treatment for all three outcomes, while the estimated within-study variance was close to zero throughout. Descriptive AIC and BIC comparisons favoured retaining the between-study heterogeneity component for all post-treatment outcomes but favoured models without this component for all follow-up outcomes (Supplementary Section S3). Prediction intervals included 0 except for follow-up interpretation bias and anxiety. Pooled

effects changed little when ρ was varied from .30 to .90 at post-treatment and when ρ and φ were jointly varied at follow-up (Table S3). Please see Fig. 2 for the caterpillar plots and Figure S9-S14 in the supplement for the forest plots for each outcome. In Fig. 2, the individual estimates (i.e., dark-grey point estimates) primarily clustered to the left of the null line across all outcomes, indicating a predominant direction of benefit favouring digital CBM-I. While the magnitude of effects varied, as shown by the S-shaped distribution, the pooled estimates (i.e., black diamonds) indicated significant improvements.

Meta-Regression Analyses: Study, Population, and Training Characteristics

Several moderator analyses, particularly at follow-up, could not be conducted because of insufficient numbers of studies within categories (see Supplementary Section S4 and Table S7 for moderator findings and Table S6 for subgroup estimates). Nine moderator tests reached nominal significance before multiplicity correction: comparator type for post-treatment interpretation bias; diagnosis for post-treatment anxiety; clinical status, diagnosis, imagery, and stimulus modality for post-treatment depression; diagnosis for follow-up anxiety; and imagery and stimulus modality for follow-up depression. Following Benjamini–Hochberg correction, no moderators remained significant except for clinical status and stimulus modality for post-treatment depression. Follow-up duration did not significantly moderate interpretation bias, anxiety, or depression.

Clinical status significantly moderated post-treatment depression effects, which were smaller in subclinical samples ($g = -0.08$, $SE = 0.06$, 95% CI $[-0.21, 0.04]$) than in clinical samples ($g = -0.33$, $SE = 0.06$, 95% CI $[-0.46, -0.21]$). The clinical-versus-subclinical contrast remained significant following Benjamini–Hochberg correction ($b = -0.25$, $SE = 0.08$, 95% CI $[-0.42, -0.08]$, $p = .0068$; BH-adjusted $p = .0338$). At follow-up, depression effects were numerically larger in clinical ($g = -0.23$, $SE = 0.06$, 95% CI $[-0.38, -0.07]$) than subclinical samples ($g = -0.15$, $SE = 0.06$, 95% CI $[-0.29, -0.01]$), but the contrast was not significant ($b = -0.08$, $SE = 0.09$, 95% CI $[-0.27, 0.11]$, $p = .383$; BH-adjusted $p = .537$). Stimulus modality also significantly moderated post-treatment depression effects, which were smaller for visual ($g = -0.10$, $SE = 0.05$, 95% CI $[-0.21, 0.01]$) than auditory delivery ($g = -0.34$, $SE = 0.05$, 95% CI $[-0.46, -0.23]$; $b = -0.24$, $SE = 0.07$, 95% CI $[-0.40, -0.08]$, $p = .0060$; BH-adjusted $p = .0338$). A similar pattern was observed at follow-up, with smaller effects for visual ($g = -0.12$, $SE = 0.06$, 95% CI $[-0.25, 0.01]$) than auditory delivery ($g = -0.30$, $SE = 0.03$, 95% CI $[-0.38, -0.21]$); however, this contrast did not remain significant following correction ($b = -0.18$, $SE = 0.07$, 95% CI $[-0.34, -0.03]$, $p = .0270$; BH-adjusted $p = .0946$).

Table 1

The effect sizes and heterogeneity among effect sizes for digital CBM-I on interpretation bias, anxiety, and depression at post-treatment and follow-up

	Model Values		Heterogeneity			
	n (k)	<i>g</i>	SE	95% CI	Var _{between} (SD)	Var _{within} (SD)
Post-Treatment Effects						
Negative Interpretation Bias	57 (324)	-0.54 *	0.05	[-0.64, -0.43]	0.10 (0.32)	0.00 (0.01)
Anxiety	49 (212)	-0.19*	0.04	[-0.27, -0.12]	0.02 (0.16)	0.00 (0.01)
Depression	37 (69)	-0.16 [†]	0.04	[-0.25, -0.07]	0.03 (0.16)	0.00 (< 0.001)
Follow-Up Effects						
Negative Interpretation Bias	16 (65)	-0.46*	0.06	[-0.58, -0.34]	0.01 (0.12)	0.00 (< 0.001)
Anxiety	26 (89)	-0.23*	0.04	[-0.31, -0.15]	0.01 (0.07)	0.00 (< 0.001)
Depression	25 (49)	-0.17 [†]	0.04	[-0.26, -0.08]	0.01 (0.08)	0.00 (< 0.001)
Note. * $p < .0001$, [†] $p < .05$, n = total number of studies, k = total number of effect sizes (as one study can include more than one comparison if they measure the same variable with different tools), <i>g</i> = effect size (Hedges' <i>g</i> , note that negative values show reduction), SE = Standard Error, 95% CI = 95% Confidence Interval, Var _{between} = Between-studies variance, Var _{within} = Within-studies variance. More negative values in Hedges' <i>g</i> indicate greater improvement favouring digital CBM-I over control.						

Note

Each panel displays the individual effect sizes contributing to the meta-analysis, sorted by magnitude to illustrate the distribution. Panels **A–C** represent results at post-treatment; panels **D–F** represent results at follow-up. The dark grey dots represent the point estimate (Hedges' *g*) for each individual effect size, while the horizontal light grey lines represent their 95% Confidence Intervals (CIs). The black diamond at the bottom of each plot denotes the overall pooled effect size and its 95% CI. The vertical dashed line indicates the point of no effect ($g = 0$); values to the left (negative) indicate a reduction in symptoms (improvement) favouring the intervention, whereas values to the right (positive) indicate an increase in symptoms (worsening). *N* represents the total number of effect sizes included in each analysis.

Discussion

Findings showed that digital CBM-I modified interpretation bias and yielded small but statistically significant symptom improvements both for anxiety and depression, with effects of similar magnitude observed at follow-up ranging from 2 to 52 weeks across outcomes. Heterogeneity mainly stemmed from between-study variance descriptively. Findings were robust to outlier exclusion and leave-one-cluster-out analyses, and no small-study effects were observed except for follow-up anxiety, although the predominance of elevated risk-of-bias ratings reduces confidence in the pooled estimates. No moderator contrast remained significant following Benjamini–Hochberg correction except for two, clinical status and stimulus modality, both for post-treatment effects on depression, warranting further investigation of digital CBM-I in depressive populations and of delivery-format effects. Overall, our estimates extended some findings from previous meta-analyses combining digital and laboratory-based delivery settings but diverged substantially from others, illustrating the need for meta-analytic examination of digital-specific evidence as a foundation for any subsequent implementation directions, as both pooled estimates and moderator results may differ across delivery settings.

Extending the findings of previous meta-analyses, effects on interpretation bias were stronger than those on symptom outcomes, in line with earlier estimates.^{14,23} The medium effect size on interpretation bias in this study ($g = -0.54$) was also consistent with the effect reported by Martinelli et al. for online CBM-I ($g = 0.44$)⁷. Though smaller in magnitude, the effects on symptoms were significant and maintained at follow-up. Effects on anxiety at post-treatment were directionally consistent with recent meta-analytic estimates, although somewhat smaller in magnitude than broader CBM-I syntheses that included non-digital delivery,⁶ and contradicted earlier estimates showing null results¹⁵ or effects reaching significance only under stress induction.^{14,23} Importantly, effects on anxiety were preserved at follow-up (up to 52 weeks), in contrast to only one study that examined this previously, which found no effect.¹⁵

Depression findings too diverged from the previous meta analyses: we observed small effect sizes, maintained at follow-up and across comparator types, unlike earlier reviews that reported null at post-treatment and/or follow-up, or comparator-contingent effects,^{6,11,14,24,25} with the exception of one recent meta-analysis by Li et al.⁸ Specifically, the size of the depression effect in our study ($g = -0.16$) was closer to Li et al.'s estimate for home-delivered CBM across modalities ($g = -0.28$) than to their much larger CBM-I subgroup estimate pooling all delivery settings ($g = -1.45$).

Several factors may explain these divergences. First, our review was intentionally restricted to digital delivery and included a more recent evidence base than even the latest previous reviews,⁶ with approximately two thirds of the depression studies and over half of anxiety studies published in 2020 or later. This may be particularly important for depression findings, as CBM-I has been increasingly adapted for depression over the past decade.^{17–19} The more favourable estimates may therefore partly reflect the maturation of intervention content and digital delivery, although this explanation could not be tested directly.

Alternative explanations include our broader inclusion criteria—grey literature and non-clinical samples—whereas some previous reviews typically restricted analyses to published studies,^{7,8,11,15,23} with some including only clinical and subclinical populations.⁶ The effects of inclusion of unpublished studies could not be tested directly due to too few unpublished studies for moderation analysis, though they did not alter the outcomes when excluded individually in leave-one-cluster-out analyses (Figures S3–S8). No significant differences between clinical and non-clinical groups for anxiety and interpretation bias were observed. On the contrary, stronger effects were observed for clinical groups for depression.

Moderation analyses highlighted multiple patterns. First, clinical samples showed stronger benefits for depression at post treatment ($g = -0.33$, 95% CI $[-0.46, -0.21]$ for clinical samples versus $g = -0.08$, 95% CI $[-0.21, 0.04]$ for subclinical samples), remaining significant even after Benjamini–Hochberg correction, although the same was not observed at follow-up. This questions the notion that CBM-I is best suited to preventive use⁷ and instead suggests its consideration and further investigation in clinical contexts for depression. Moreover, the effects were stronger specifically for depression diagnosis compared to other diagnoses ($g = -0.37$, 95% CI $[-0.63, -0.11]$ for depression; $g = -0.22$, 95% CI $[-0.41, -0.04]$ for general anxiety; and $g = -0.07$, 95% CI $[-0.18, 0.03]$ for other diagnoses), but the moderation was not significant after correction. Another significant moderation for depression was observed for stimulus modality both at post-treatment and follow-up, where auditory formats produced more than twice the effect of visual delivery, though only the post-treatment effects survived correction. Collectively, these findings warrant further investigation of the therapeutic potential of digital CBM-I for depression in clinical populations, as well as auditory delivery as a potentially more effective alternative or addition to text-based formats, as the current evidence base suggests but does not conclude strongly on these points.

Several candidate moderators, including comparator, diagnosis, and imagery, produced nominally significant contrasts that did not survive correction. First, comparator type was only significant for post-treatment interpretation bias (neutral CBM-I > sham comparators), consistent with most previous reviews,^{7,8,23}. The contrast with Fodor et al.'s⁶ findings for depression challenges the suggestion that CBM-I effects on depression depend on comparator type. Second, there were only nominally significant moderation effects by imagery (imagery > no imagery) that did not survive correction, which contradicts earlier significant findings.²⁶ One possible explanation could be implementation differences: engagement in imagery practice during the intervention can be monitored and encouraged in lab-based studies but less so in remote digital delivery, and no trial has yet measured actual imagery adherence. Third, diagnosis was a significant moderator for anxiety before correction at both post-treatment and follow-up (samples with anxiety diagnoses > samples with diagnoses classified as other), yet the heterogeneous composition of the 'other diagnosis' category limits the clinical interpretability of this contrast. No evidence of moderation was found for dose, items per session, delivery mode, stimulus type, or age. The absence of moderating effects of dose aligns with recent meta-analyses^{6,7,23} but contrasts with earlier reports.^{14,26} However, a granular dose-response examination was not possible here. Finally, we found no evidence that effect size varied with follow-up duration, though follow-ups

were relatively short (range = 2–52 weeks). Overall, some non-significant findings here highlight the need for evidence specific to digital delivery, as some significant moderation effects in lab-based studies (e.g., comparator, imagery) appear not to generalise to digital delivery in the current evidence base.

Modified Egger's tests indicated small-study effects only for follow-up anxiety effects, with funnel plots showing mixed findings. Sensitivity analyses suggested that either large publication bias parameters would be needed to nullify results, or no such finite selection ratio was able to nullify results,²⁷ and leave-one-out analyses confirmed no single study drove effects, both indicating robustness. Main meta-analytic findings remained robust even after the outlier exclusion and statistical correction. Collectively, these analyses indicate that the findings were robust to the specific sensitivity checks conducted; however, the predominance of high-risk studies and evidence of small-study effects for follow-up anxiety mean that methodological bias and selective reporting cannot be ruled out.

Several limitations should be noted. Some moderators (e.g., publication status) could not be tested due to sparse data. Also, meta-analyses were conducted on the basis of outcome (e.g., depression) and diagnosis-specific effects were investigated in the moderator analyses (e.g., effects on depression for clinically depressed samples). Thus, effects on 'anxiety' and 'depression' should be interpreted dimensionally, rather than as evidence of efficacy for particular diagnoses. Furthermore, follow-up periods were relatively short (two to 52 weeks), limiting conclusions on longer-term efficacy. Additionally, contrary to post-treatment analyses, follow-up analyses included fewer studies (16, 26, and 25 for bias, anxiety, and depression, respectively) than the number typically recommended for highly stable three-level variance estimation ($\approx 40-50$),²² and therefore these findings should be interpreted with caution. Finally, the results should be considered in light of possible decline effects, whereby intervention effect sizes attenuate as evidence accumulates and interventions are independently replicated, particularly given the recency of much of the digital CBM-I literature.⁷

Despite these caveats, this study provides the most comprehensive synthesis of digital CBM-I to date, arriving at a timely moment given recent calls for large-scale implementation.³ Findings demonstrate that digital CBM-I modifies interpretation bias and yields small but significant symptom improvements both for anxiety and depression, with the effects preserved at follow-up and larger effects observed for depression in clinical samples. These results contrast with several earlier null or less consistent estimates and support continued evaluation of digital CBM-I as a low-intensity intervention for anxiety and depression. More broadly, the divergence between our findings and those of meta-analyses combining digital and laboratory delivery underscores the importance of synthesising evidence specific to digital delivery and examining moderators within the intended delivery context.

For future research with policy relevance, five priorities emerge. First, trials with longer follow-up periods are needed to define the durability of effects. Second, more adequately powered studies in diagnostically defined depressive samples are needed to support reliable diagnosis-specific analyses. Third, digital CBM-I appears to produce stronger effects on depression in clinical samples at post-treatment, warranting exploration of its clinical potential further. Fourth, future trials and evidence syntheses should

extend beyond standardised mean effects by incorporating clinically meaningful improvement, responder outcomes, safety, engagement, and health economic evaluation.²⁸ This is particularly important for digital interventions as even slight but clinically meaningful improvements on targeted symptoms could be of value if delivered at low cost. Fifth, efficacy evaluation of digital CBM-I should move beyond diagnosis-level symptom totals and remission outcomes to examine symptom-level effects and precision. Such analyses could identify which symptom profiles are most responsive and inform the development of more personalised intervention delivery. Addressing these priorities would determine whether statistically detectable benefits of digital CBM-I translate into clinically meaningful, safe, cost-effective, and implementable benefits in routine care.

Methods

Eligibility Criteria

For this review, digital CBM-I was defined as CBM-I delivered through a web-based platform, mobile application, or other digital interface in a format that participants could complete without researcher-led administration. Studies in which CBM-I was delivered exclusively during researcher-led laboratory sessions were therefore excluded, even when stimulus presentation itself was computerised.

Inclusion criteria were (1) studies administering digital CBM-I with or without a control group and randomization, (2) studies measuring at least one of the following outcomes: interpretation bias, anxiety, or depression (continuous or categorical; state measures of anxiety, affect, or mood were ineligible as trait, rather than momentary, changes were of interest), (3) any samples, including healthy, subclinical, and clinical, and (4) published or unpublished reports, including pre-prints, dissertations, and conference papers.

Exclusion criteria were (1) animal studies, (2) publications or reports not in English, (3) studies involving participants with neurological diseases and/or injuries, and/or learning or developmental disorders, (4) studies in which participants in the CBM-I condition (but not the comparator condition/s) received additional interventions at or around the same time (i.e., confounding the results), and (5) studies delivering CBM-I exclusively in researcher-led sessions (e.g., lab-based studies).

Search strategy

The search, conducted in June 2026, covered PsycINFO, MEDLINE, Embase (all via OVID), SCOPUS, Web of Science, and unpublished sources (PsyArXiv and ProQuest Dissertations) with no time and language limitations applied. The search strategy combined terms for CBM-I and digital delivery and excluded unrelated uses. The search terms therefore were (“Cognitive bias modification for interpretation” OR “Interpretation training” OR CBM-E OR “Interpretation bias modification” OR CBM-I OR “Interpretation bias training” OR “modify* interpretation”) AND (digital OR online OR eHealth OR mHealth OR online-delivered

OR “mobile app” OR smartphone OR “mobile phone app” OR internet) NOT (“interpreter training” OR mammographic)

Study Selection

Search results were uploaded to Rayyan for screening. Screening was conducted at title/abstract and full-text stages by two independent blind reviewers (XW, MS), with KK supervising the process, preparing screening and coding guidelines, training reviewers, and resolving conflicts through discussion. Conflicts that could not be resolved by KK were escalated to the senior authors (JY, JE), with consensus reached for all decisions. When required data were missing, corresponding authors were contacted to request the relevant data or information.

Two independent coders (XW, MS) extracted study data across several domains, with conflicts resolved by a third reviewer (KK). Study identifiers included authors, title, publication year, journal, and country. Outcomes comprised all reported measures of interpretation bias, depressive symptoms, and anxiety, with means and standard deviations extracted for intervention and, where applicable, control.

Candidate moderators were also coded in the following categories. Study characteristics covered publication status, risk of bias, randomisation status. Sample characteristics included group size, age group (adult/minor, or both), mean age, gender distribution, clinical status (clinical, subclinical, or healthy), and DSM-5 and/or ICD-11 diagnoses where available. Training and design features included imagery use, training paradigm (ambiguous scenarios/other formats), delivery mode (mobile app/web), dose (sessions × items), stimulus type (verbal/non-verbal), stimulus modality (visual/auditory), and comparator condition (neutral CBM-I, negative CBM-I, unresolved ambiguity, or waiting list/treatment as usual). Authors were contacted for missing and unpublished data. For each follow-up effect size, assessment time was coded as the number of weeks from baseline to the follow-up assessment.

Risk of Bias Assessment

Risk of bias was assessed by two independent researchers (ZJ, KL) using the Cochrane Risk of Bias 2 (RoB2) tool for randomized controlled trials and the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool for non-randomized studies (only included in the supplementary within-group meta-analyses, not in the main between-group meta-analyses), following Cochrane Handbook guidance.²⁹ Discrepancies were resolved through discussion between reviewers, with unresolved cases adjudicated by a third reviewer (KK). RoB2 studies were classified as ‘high’, ‘moderate/some concerns’, or ‘low risk’, and ROBINS-I studies had also ‘serious’ and ‘critical’ as higher risk categories.²⁹

Data Analysis

The primary meta-analyses reported in the manuscript synthesised between-group effects (i.e., CBM-I vs. control groups at post-treatment/follow-up). For the primary meta-analyses, effect sizes (ESs) were calculated as between-group standardized mean differences (SMD), using means, standard deviations (SD), and sample sizes. The SMDs were corrected for small sample bias (Hedges's g). 95% CI was calculated for all effect sizes. Even though non-randomised studies were not excluded from the systematic review, the primary meta-analyses reported in the manuscript only included between-group effects from randomised studies (there were no studies with a control group that were non-randomised). The prespecified significance level was $\alpha = .05$, and all statistical tests were two-sided. Intention-to-treat data were used when available, and if not, available-case/completer endpoint statistics and their corresponding analysed sample sizes were used. Where there was a secondary use of the same data, only the original study was included in effect size calculations. A three-level random-effects meta-analytic model with restricted maximum likelihood estimation was employed to allow the inclusion of multiple effect sizes per study.²⁰ Separate within-group analyses of baseline-to-post-treatment changes in CBM-I groups were conducted, as pre-registered, to examine whether the between-group findings reflected improvement in the intervention group rather than deterioration in the control group; these analyses are reported in the supplementary materials (see Table S4). Supplementary within-group effects were calculated as bias-corrected standardized mean changes using post-treatment minus baseline means, the pooled pre/post SD, the paired post-treatment sample size, and an assumed pre-post correlation of $r = .50$ (`metafor::escalc, measure = "SMCRP", correct = TRUE`). Within-group follow-up analyses were not conducted because they were not pre-registered. Post-treatment and follow-up effects were analysed separately for interpretation bias, anxiety, and depression, resulting in six meta-analyses (2 time points x 3 outcomes). Benjamini-Hochberg adjustment was applied post-hoc across the six primary pooled-effect tests for Type I error correction.³⁰ Negative values indicate greater reductions/improvements in negative interpretation bias, depression, or anxiety.

Primary models first included all eligible between-group effect sizes. Separately within each outcome and time point, raw Hedges' g values lying more than 2.5 unweighted sample SDs below or above the unweighted arithmetic mean were classified as potential outliers; models excluding these effects were fitted as sensitivity analyses.³¹ Heterogeneity was examined using the estimated variance components (σ^2) for variability among effect sizes within studies (level 2) and variability between studies (level 3), together with comparisons against models in which each variance component was fixed to zero.²⁰ Sampling error was incorporated directly through the complete variance-covariance matrices in all models. Sampling variance-covariance matrices constructed using `vcalc()` to account for dependencies arising from multiple outcomes and shared intervention or control arms included positive-semidefinite, singular study-specific blocks.³² Consequently, conventional generalized multilevel I^2 statistics could not be calculated. In some multi-arm studies, complete connected sets of contrasts introduced exact linear dependencies, rendering the sampling variance-covariance matrix rank-deficient. Because the conventional generalized I^2 calculation for correlated sampling errors requires inversion of this matrix, level-specific I^2 statistics were not calculated. Heterogeneity was therefore summarised using the REML variance components, the distribution of estimated heterogeneity across the within- and between-study

levels, prediction intervals, and descriptive AIC and BIC comparisons. Caterpillar plots were generated to visualise the distribution of effect sizes for each outcome. These plots are well-suited for meta-analyses containing a substantial number of effect sizes, as they provide a clear overview of the distribution, consistency, and potential outliers that standard forest plots cannot accommodate at this scale. Meta-regression analyses tested the pre-registered moderator hypotheses. Categorical moderators were tested formally only when at least two retained categories each contained a minimum of five independent studies with Benjamini–Hochberg adjustment applied separately within each outcome-by-timepoint family. Randomisation was not examined as a moderator because the moderator analyses were restricted to controlled between-group comparisons, which only included randomised trials. Subgroup-specific contrasts were treated descriptively and were not entered as additional multiplicity tests. Reporting bias and small-study effects were examined using funnel plots, including funnel plots, modified Egger’s test,³³ sensitivity analysis,²⁷ and leave-one-cluster-out analysis. Risk of bias was used in moderator analyses to examine any systematic differences across risk levels, though only in post-treatment moderator analyses and not in follow-up moderator analyses, as the risk-of-bias judgements used in the review concerned post-treatment outcomes and were not separately assessed for follow-up outcomes.

Analyses were conducted in R (version 4.4.3), using *meta*³⁴ (version 8.2-0) for effect-size calculation, *metafor*³² (version 4.8-0) for multilevel models, *clubSandwich*³⁵ (version 0.6.1) for cluster-robust inference, and *PublicationBias*²⁷ (version 2.4.0) for significance-selection sensitivity analyses. The *rma.mv* function was used for the three-level meta-analyses. The *coef_test()* function from the *clubSandwich* package³⁵ was employed to apply cluster-robust standard errors with Satterthwaite degrees of freedom to account for dependent sampling errors of the effect sizes coming from the same study. For the primary between-group analyses, sampling variance–covariance matrices were constructed using *vcalc()* to account for dependencies arising from multiple outcomes and shared intervention or control arms, assuming ρ (rho) = .50 between different outcomes measured concurrently, with sensitivity analyses re-fitting the post-treatment models using ρ = .30, .50, .70, and .90. Follow-up models additionally accounted for repeated follow-up measurements using a temporally decaying correlation structure with a primary one-month autocorrelation of ϕ (phi) = .80 and tested models across all combinations of ρ = .30, .50, .70, and .90 and ϕ = .30, .50, .70, .80, and .90 for sensitivity analyses. The study was pre-registered (PROSPERO ID: CRD42023489316) and reported following PRISMA guidelines. The PRISMA 2020 checklist,³⁶ the R script, HTML outputs, and dataset are provided on OSF (https://osf.io/h9wt5/overview?view_only=79e06997654d432c86721dadd9d450ab). The funders of the study had no role in study design, data collection, analysis, interpretation, or writing of the report.

Declarations

Acknowledgments

The authors would like to thank **Rayan Taher** and **Ava Cooney** for their contributions in double-checking a portion of inclusion decisions and data extraction. The authors used ChatGPT (OpenAI, GPT-5.6) to

assist with improving the readability and language of the work. The AI tool was not used for study design, data analysis, or interpretation. All content was reviewed and approved by the authors, who take full responsibility for the manuscript. KAK is supported by a doctoral scholarship from the Ministry of National Education of Türkiye. JY and KAK were supported by a grant from the Medical Research Council Impact Acceleration Account (MRC IAA) 2021 Kings College London (MR/X502923/1). JY is part-funded by the NIHR Biomedical Research Centre hosted at South London and Maudsley NHS Foundation Trust in partnership with King's College London. JE is supported by a grant from the Research Foundation Flanders grant (1202122N). The funders had no role in the design, analysis, interpretation, or reporting of this study.

Author Contributions

KAK, JY, JE, and SS were responsible for study concept and design. KAK planned and supervised the screening and data extraction processes, which were conducted by XW and MS. ZJ and KL were responsible for risk of bias assessment under the supervision of KAK.

KAK was responsible for acquisition, statistical analysis, and interpretation of data, with input from JE and JY. KAK was responsible for drafting of the manuscript. JY, JE, and SS were responsible for critical revision of the manuscript. JY and JE were responsible for study supervision. DS supervised methodological decisions and their statistical applications. All authors reviewed and approved the final version of the manuscript.

Competing Interests

All authors declare no competing interests.

Data Availability

The datasets generated and analysed are available on the Open Science Framework at https://osf.io/h9wt5/overview?view_only=79e06997654d432c86721dadd9d450ab.

Code Availability

The R scripts used in the analysis are available on the Open Science Framework at https://osf.io/h9wt5/overview?view_only=79e06997654d432c86721dadd9d450ab.

References

1. World Health Organization. *Depression and Other Common Mental Disorders: Global Health Estimates*. (World Health Organization, Geneva, Switzerland, 2017).
2. Layard, R. & Clark, D. M. Why More Psychological Therapy Would Cost Nothing. *Front. Psychol.* 6, (2015).
3. Vrijzen, J. N. et al. Towards implementation of cognitive bias modification in mental health care: State of the science, best practices, and ways forward. *Behav. Res. Ther.* 179, 104557 (2024).

4. Beck, A. T. *Cognitive Therapy and the Emotional Disorders*. (International Universities Press, 1976).
5. Mathews, A. & MacLeod, C. Cognitive Vulnerability to Emotional Disorders. *Annu. Rev. Clin. Psychol.* 1, 167–195 (2005).
6. Fodor, L. A. et al. Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: a systematic review and network meta-analysis. *Lancet Psychiatry* 7, 506–514 (2020).
7. Martinelli, A., Grüll, J. & Baum, C. Attention and interpretation cognitive bias change: A systematic review and meta-analysis of bias modification paradigms. *Behav. Res. Ther.* 157, 104180 (2022).
8. Li, J., Ma, H., Yang, H., Yu, H. & Zhang, N. Cognitive bias modification for adult's depression: A systematic review and meta-analysis. *Front. Psychol.* 13, 968638 (2023).
9. Steinman, S. A. et al. Which Variations of a Brief Cognitive Bias Modification Session for Interpretations Lead to the Strongest Effects? *Cogn. Ther. Res.* 45, 367–382 (2021).
10. Salemink, E., Kindt, M., Rienties, H. & Van Den Hout, M. Internet-based Cognitive Bias Modification of Interpretations in patients with anxiety disorders: A randomised controlled trial. *J. Behav. Ther. Exp. Psychiatry* 45, 186–195 (2014).
11. Sicouri, G. et al. Cognitive bias modification of interpretations for anxiety and depression in children and adolescents: A meta-analysis. *JCPP Adv.* 4, e12207 (2024).
12. Antognelli, S. L., Sharrock, M. J. & Newby, J. M. A randomised controlled trial of computerised interpretation bias modification for health anxiety. *J. Behav. Ther. Exp. Psychiatry* 66, 101518 (2020).
13. Larrazabal, M. A. et al. Online cognitive bias modification for interpretation to reduce anxious thinking during the COVID-19 pandemic. *Behav. Res. Ther.* 173, 104463 (2024).
14. Hallion, L. S. & Ruscio, A. M. A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychol. Bull.* 137, 940–958 (2011).
15. Kampmann, I. L., Emmelkamp, P. M. G. & Morina, N. Meta-analysis of technology-assisted interventions for social anxiety disorder. *J. Anxiety Disord.* 42, 71–84 (2016).
16. Williams, A. D. et al. Positive imagery cognitive bias modification (CBM) and internet-based cognitive behavioral therapy (iCBT): A randomized controlled trial. *J. Affect. Disord.* 178, 131–141 (2015).
17. Lee, J.-S. et al. How can we enhance cognitive bias modification techniques? The effects of prospective cognition. *J. Behav. Ther. Exp. Psychiatry* 49, 120–127 (2015).
18. Lester, K. J., Mathews, A., Davison, P. S., Burgess, J. L. & Yiend, J. Modifying cognitive errors promotes cognitive well being: A new approach to bias modification. *J. Behav. Ther. Exp. Psychiatry* 42, 298–308 (2011).
19. Yiend, J. et al. Modifying Interpretation in a Clinically Depressed Sample Using 'Cognitive Bias Modification-Errors': A Double Blind Randomised Controlled Trial. *Cogn. Ther. Res.* 38, 146–159 (2014).
20. Cheung, M. W.-L. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol. Methods* 19, 211–229 (2014).

21. Hirsch, C. R. *et al.* Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. *J. Consult. Clin. Psychol.* 86, 1017–1030 (2018).
22. Moeyaert, M. *et al.* Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *Int. J. Soc. Res. Methodol.* 20, 559–572 (2017).
23. Krebs, G. *et al.* Research Review: Cognitive bias modification of interpretations in youth and its effect on anxiety: a meta-analysis. *J. Child Psychol. Psychiatry* 59, 831–844 (2018).
24. Cristea, I. A., Kok, R. N. & Cuijpers, P. Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *Br. J. Psychiatry* 206, 7–16 (2015).
25. Liu, H., Li, X., Han, B. & Liu, X. Effects of cognitive bias modification on social anxiety: A meta-analysis. *PLOS ONE* 12, e0175107 (2017).
26. Menne-Lothmann, C. *et al.* How to Boost Positive Interpretations? A Meta-Analysis of the Effectiveness of Cognitive Bias Modification for Interpretation. *PLoS ONE* 9, e100925 (2014).
27. Mathur, M. B. & VanderWeele, T. J. Sensitivity Analysis for Publication Bias in Meta-Analyses. *J. R. Stat. Soc. Ser. C Appl. Stat.* 69, 1091–1119 (2020).
28. Bauer-Staeb, C. *et al.* Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials. *J. Clin. Epidemiol.* 137, 200–208 (2021).
29. Higgins, J. *et al.* *Cochrane Handbook for Systematic Reviews of Interventions Version 6.5.* (2024).
30. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300 (1995).
31. Lipsey, M. W. & Wilson, D. B. *Practical Meta-Analysis.* ix, 247 (Sage Publications, Inc, Thousand Oaks, CA, US, 2001).
32. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* 36, (2010).
33. Pustejovsky, J. E. & Rodgers, M. A. Testing for funnel plot asymmetry of standardized mean differences. *Res. Synth. Methods* 10, 57–71 (2019).
34. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* 22, 153–160 (2019).
35. Pustejovsky, J. E., Pekofsky, S. & Zhang, J. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. (2025).
36. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* 88, 105906 (2021).

Figures

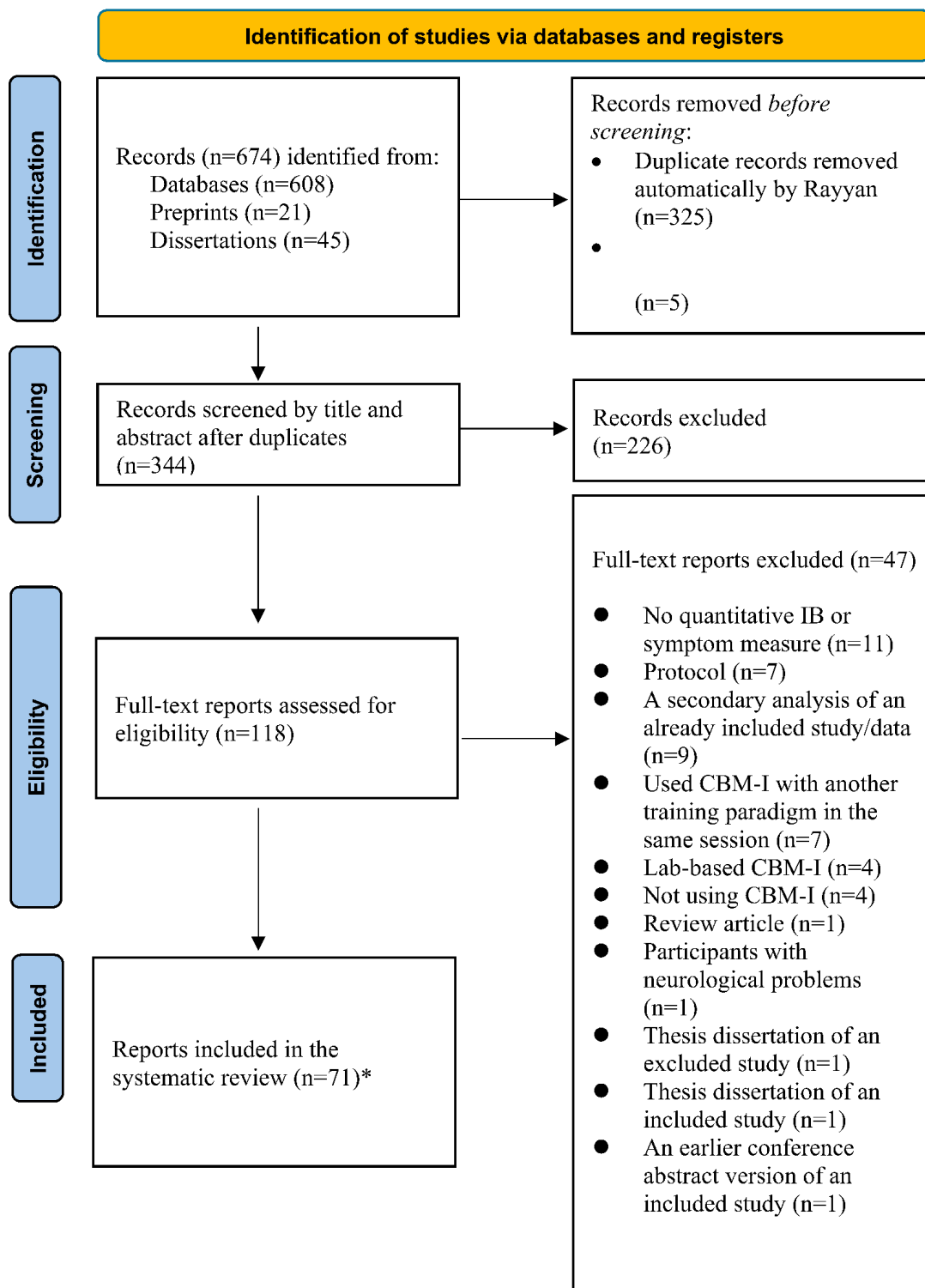


Figure 1

PRISMA flow diagram describing the identification, screening, and inclusion of the relevant reports in the systematic review.

*The number of studies included in the systematic review is 74, as some papers report multiple independent experiments.

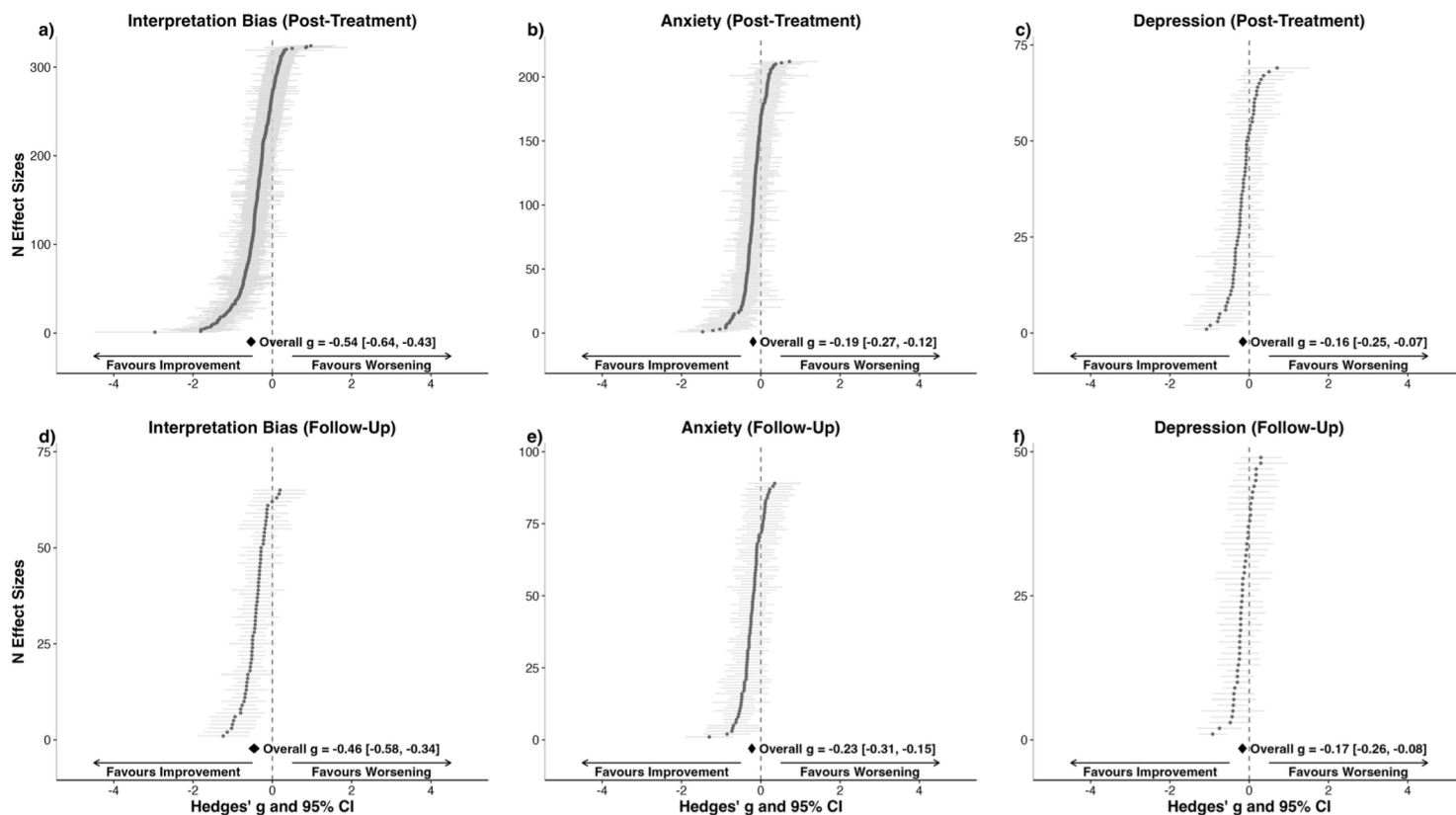


Figure 2

Distribution of effect sizes for interpretation bias, anxiety, and depression at post-treatment and follow-up

Note. Each panel displays the individual effect sizes contributing to the meta-analysis, sorted by magnitude to illustrate the distribution. Panels **A–C** represent results at post-treatment; panels **D–F** represent results at follow-up. The dark grey dots represent the point estimate (Hedges' g) for each individual effect size, while the horizontal light grey lines represent their 95% Confidence Intervals (CIs). The black diamond at the bottom of each plot denotes the overall pooled effect size and its 95% CI. The vertical dashed line indicates the point of no effect ($g = 0$); values to the left (negative) indicate a reduction in symptoms (improvement) favouring the intervention, whereas values to the right (positive) indicate an increase in symptoms (worsening). N represents the total number of effect sizes included in each analysis.

Supplementary Files

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

- [DigitalCBMISRMAV1.0Supplement.pdf](#)