

Bartoš, František; Godmann, Henrik R.

Working Paper

A Comment on "A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders"

I4R Discussion Paper Series, No. 213

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Bartoš, František; Godmann, Henrik R. (2025) : A Comment on "A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders", I4R Discussion Paper Series, No. 213, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/313186>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

No. 213

I4R DISCUSSION PAPER SERIES

A Comment on “A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders”

František Bartoš

Henrik R. Godmann

March 2025

I4R DISCUSSION PAPER SERIES

I4R DP No. 213

A Comment on “A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders”

František Bartoš¹, Henrik R. Godmann²

¹Charles University, Prague/Czech Republic

²University of Amsterdam/The Netherlands

MARCH 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

A comment on “A systematic review and meta-analysis of transdiagnostic cognitive behavioural therapies for emotional disorders”*

František Bartoš and Henrik R. Godmann

December 12, 2024

Abstract

Schaeuffele et al. (2024) examined the effect of Transdiagnostic Cognitive Behavioural Psychotherapy (TD-CBT) on emotional disorders through a meta-analysis of 53 studies involving 6,705 participants. Their main findings indicated that TD-CBT has larger treatment effects on depression, $g = 0.74$, 95% CI (0.57, 0.92), $p < 0.001$, and anxiety, $g = 0.77$, 95% CI (0.56, 0.97), $p < 0.001$, than controls. We replicated the data extraction of summary information from a subsample of the original studies with only minor deviations and we successfully computationally reproduced the main claims of the paper using the original data. However, robustness analyses adjusting for publication bias led to substantially different conclusions. In contrast to the original authors, we found weak evidence for the absence of the overall treatment effect of TD-CBT on both depression and anxiety in most of the specified models. Although additional sensitivity analyses could not completely rule out that the observed differences are partially results of small-study effects, further examination of between-study heterogeneity did not reveal consistent evidence of the benefit of TD-CBT when compared to any of the comparisons groups.

KEYWORDS: Depression, Anxiety, Publication Bias, Bayesian, Heterogeneity

*Authors: František Bartoš: Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague. E-mail: f.bartos96@gmail.com (corresponding author). Henrik R. Godmann: Department of Psychological Methods, University of Amsterdam. E-mail: hrgodmann@gmail.com. The authors declare that they do not have a conflict of interest. František Bartoš acknowledges support from the Czech Science Foundation (grant #23-05227M).

1 Introduction: Example

Schaeuffele et al. (2024) conducted a systematic review and meta-analysis of Transdiagnostic Cognitive Behavioral Therapies (TD-CBT) as a treatment for depression and anxiety. Two independent reviewers screened studies and extracted data from randomized controlled trials (RCTs) identified via database searches in PubMed, MEDLINE, PsycINFO, Google Scholar, medRxiv, and OSF spanning from 2000 to 2023. In their main analysis, Schaeuffele et al. (2024) examined the effectiveness of TD-CBT in comparison to controls at post-treatment in reducing symptoms of depression and anxiety using random-effects meta-analytic models. Schaeuffele et al. (2024) found large overall effects of TD-CBT on the decrease in both depression, $g = 0.74$, 95% CI (0.57, 0.92), $p < 0.001$, and anxiety, $g = 0.77$, 95% CI (0.56, 0.97), $p < 0.001$, symptoms (abstract, and p. 501).

In this report prepared for the Institute for Replication in collaboration with Nature Human Behaviour (Brodeur et al. 2024), we first replicated the data extraction procedure for obtaining summary information from the original studies with computed effect sizes larger than $d = 3$ (two for depression and four for anxiety; we selected those estimates since very large effect sizes are more likely a result of a coding error). We found only minor deviations (summarized in Table 1) which we attributed to poor reporting in the original studies, which would not meaningfully affect the original results. We proceeded to successfully reproduce the main analyses and additional results reported in Table 2 in Schaeuffele et al. (2024) using the original and publicly shared R code and data (<https://osf.io/ta4fg/>). The code was complete and did not require additional adjustments.

Second, we performed a series of planned robustness analyses of the main findings to examine whether the main findings are robust to a) publication bias adjustment, b) publication bias adjustment and adjustment for sample overlap due to the common comparison groups, and c) publication bias adjustment, adjustment for sample overlap due to the common comparison groups, and exclusion of extreme effect sizes. In all analyses, we found that the overall effect of TD-CBT disappears. Furthermore, results often suggest weak evidence for an *absence* of the overall effect of TD-CBT on reducing symptoms of depression and anxiety, extreme evidence for the presence of heterogeneity, and strong evidence for the presence of publication

bias.

Third, we conducted a series of unplanned sensitivity analyses. We examined whether the conclusions of the planned robustness analyses held when omitting adjustment for publication bias that could be driven by small-study effects. We further examined heterogeneity between studies in different comparison groups using meta-regression models and subgroup analyses. The sensitivity analyses verified the weak evidence for the absence of the treatment effect of TD-CBT on depression held even when using publication bias adjustment robust to small-study effects; however, the sensitivity analyses of the effect of TD-CBT on anxiety showed moderate evidence for the presence of the effect when using publication bias adjustment robust to small-study effects. Finally, the examination of between-study heterogeneity due to different comparison groups showed no consistent benefit of TD-CBT when compared to any of the comparison groups in the sensitivity analyses. We found moderate or even strong evidence in favor of TD-CBT when compared to either treatment as usual (TAU) or wait-list for both depression and anxiety when performing a publication bias adjusted meta-regression; however, publication bias subgroup analyses resulted in inconsistent conclusions, suggesting potential differences in heterogeneity and publication bias among the comparison groups.

2 Computational Reproducibility

We manually replicated the data extraction procedure of summary information required for calculating the effect size for all effect size estimates larger than $d = 3$. This selection resulted in extracting summary information for two effect size estimates of depression and four effect size estimates of anxiety. Table 1 summarizes the re-extracted vs. original summary information showing only minor differences in three effect size estimates. We would attribute these minor differences to unclear reporting in the original studies, and we do not believe that they would impact the original conclusions. Therefore, we conduct all subsequent analyses with the original data.

We proceeded by reproducing the main findings and results from Table 2 in Schaeuffele et al. (2024) based on the dataset and R scripts publicly shared by the original authors at OSF (<https://osf.io/ta4fg/>). The materials were complete

and allowed us to computationally reproduce the results from the raw data.

3 Robustness Reproduction

Schaeuffele et al. (2024) based their main findings and additional results in Table 2 on random-effects meta-analyses estimated using restricted maximum likelihood via the `metafor` R package (Viechtbauer 2010). While Schaeuffele et al. (2024) reported publication bias analyses indicating the presence of publication bias “Statistical analyses indicated asymmetry of funnel plots for controlled effects (Kendall’s tau = 0.36 – 0.38, $P < 0.001$; Egger’s test $Z = 6.89 - 7.45$, $P < 0.001$). Funnel plots, with and without trim and fill method, for the controlled effects for anxiety and depression (post-treatment) are provided in Supplementary Figs. 20 and 21.” (p. 503), they do not report or further discuss publication bias adjusted effect size estimates in the main manuscript.¹ We recreated a funnel plot of the effect sizes of TD-CBT on both depression and anxiety in Figure 1. The funnel plot with 90%, 95%, and 99% prediction intervals centered at 0 is indicative of strong publication bias.

We performed three planned robustness analyses to assess the impact of publication bias on the main findings. We used robust Bayesian meta-analysis (RoBMA, Maier et al. 2023) implemented in the `RoBMA` R package (Bartoš and Maier 2020). Specifically, we used `RoBMAPSMA` (Bartoš et al. 2022) which specifies 36 competing models corresponding to the different assumptions about the data; the presence vs. absence of the effect, heterogeneity, and publication bias (no bias vs. six selection models, Vevea and Hedges 1995, and two PET and PEESE style models, Stanley and Doucouliagos 2014). RoBMA employs Bayesian model averaging to combine the specified models based on their prior predictive performance and provides model-averaged effect size estimates and inclusion Bayes factors (e.g., Hoeting et al. 1999, Hinne et al. 2020). Inclusion Bayes factors are used to quantify the evidence in favor of the effect (BF_{10}), heterogeneity (BF_{rf}), and publication bias (BF_{pb}). As a rule of thumb, Bayes factors between 1 and 3 are regarded as weak evidence, Bayes factors

¹Further examination of the provided code reveals that the originally performed publication bias adjustment results in slightly smaller, yet statistically significant meta-analytic effects. However, both the trim and fill and Egger’s tests were repeatedly found to provide insufficient publication bias adjustment (Carter et al. 2019, Kvarven et al. 2020, Hong and Reed 2020).

between 3 and 10 are regarded as moderate evidence, and Bayes factors larger than 10 are regarded as strong evidence (e.g., Appendix I, [Jeffreys 1939](#); p. 105, [Lee and Wagenmakers 2013](#)). To quantify evidence for the null (absence of an effect), the Bayes factor is inverted (i.e., $BF_{01} = 1/BF_{10}$).

First, we used the default RoBMA_{PSMA} model as specified in ([Bartoš et al. 2022](#)). Second, we adjusted the results of the first analysis by accounting for the sample overlap in estimates from the same study (i.e., studies providing multiple treatment effects of TD-CBT by comparing to multiple control groups) by weighting the likelihood of the RoBMA_{PSMA} model using an approach inspired by [Bom and Rachinger \(2020\)](#). Third, we adjusted the results of the second analysis by removing effect size estimates larger than $d = 3$ (three for depression and four for anxiety).

Then, we performed additional unplanned sensitivity analyses. We examined whether results from the first planned robustness analysis (i.e., RoBMA_{PSMA} model) hold when omitting either the PET and PEESE style models or selection models from the publication bias adjustment component of RoBMA_{PSMA}. PET and PEESE style models (i.e., adjusting for the relationship between effect sizes and standard errors or sampling variances) can potentially over-adjust the effect size estimate in the presence of small-study effects ([Lau et al. 2006](#)). As such, removing the PET and PEESE style adjustment allows us to circumvent this potential issue. Finally, we further examined the substantial between-study heterogeneity by extending the RoBMA_{PSMA} model into a robust Bayesian meta-regression with the different comparison groups (RoBMA-regression, [Bartoš et al. 2024](#)). We also performed subgroup analyses with RoBMA_{PSMA} to check that the results of RoBMA-regression hold if the effects of TD-CBT differ in heterogeneity or publication bias across the comparison groups.

All results and analysis scripts are accessible at OSF (<https://osf.io/7fyu9/>).

3.1 Results of Robustness Analyses

All three planned robustness analyses showed weak evidence in favor of the absence of the overall effect of TD-CBT on depression (all BF_{10} between 0.34 and 0.38) or anxiety (all BF_{10} between 0.42 and 0.61). This conclusion is contrary to the original analyses, which rejected the null hypotheses of no effect with $p < 0.001$ in

each instance. Furthermore, the publication bias-adjusted model-averaged overall effect size estimates of TD-CBT reduced significantly, from 0.74 to 0.05-0.07 for depression and from 0.77 to 0.06-0.12 for anxiety; see Table 2 for a summary of effect size estimates, tests, and confidence/credible intervals. The difference in the conclusions results from the strong to extreme evidence of publication bias in both the depression (all BF_{pb} between 45 and 265) and anxiety (all BF_{pb} between 10 and 93) robustness analyses. Furthermore, all analyses showed substantial between-study heterogeneity, indicating that the overall treatment effect might not be representative of different settings or comparison groups.

3.2 Results of Sensitivity Analyses

Table 3 summarizes results of sensitivity analyses of the publication bias adjustment with respect to potential small-study effects. While the overall effect of TD-CBT on depression does not substantially change regardless of whether only selection models or only PET and PEESE publication bias adjustment is used, we find moderate evidence of the overall effect of TD-CBT on anxiety if only selection models publication bias adjustment is used. This suggests that some of the differences in the overall effect of TD-CBT on anxiety between the original analysis and the robustness analyses might be partially due to small-study effects.

Table 4 compares results of the original subgroup analysis to RoBMA-regressions and subgroup analyses with $RoBMA_{PSMA}$ when examining the moderating effect of the type of comparison group. Importantly, RoBMA-regression finds extreme evidence for the moderation in both the depression ($BF_{mod} = 5093$) and anxiety ($BF_{mod} = 23013$). The results of the original subgroup analyses, RoBMA-regression, and $RoBMA_{PSMA}$ subgroup analyses consistently show a lack of benefit when comparing TD-CBT to disorder-specific CBT (DS-CBT) in either depression or anxiety (non-significant p-values and weak or moderate evidence for H_0). Comparison of TD-CBT to the treatment as usual (TAU) is less consistent as the original analyses reject the null hypothesis of no effect with $p < 0.001$ for both the depression and anxiety, while RoBMA-regression shows barely moderate evidence in favor of the alternative hypotheses and subgroup analyses with $RoBMA_{PSMA}$ show weak evidence for the alternative hypotheses. The model-averaged effect size estimates adjusted

for publication bias also decrease by two- to three-fold in comparison to the original results. Comparison of TD-CBT to other comparison groups was inconsistent in the case of depression as the original analysis rejected the null hypothesis with $p < 0.001$, while the RoBMA-regression and RoBMA_{PSMA} subgroup analysis did not show even weak evidence for the presence of a positive effect (RoBMA-regression most likely overcorrects for publication bias when it suggests weak evidence for the presence of a negative effect). Comparison of TD-CBT to other comparison groups was however consistent in the case of anxiety as neither the original nor the sensitivity analyses show evidence for a positive treatment effect (RoBMA-regression most likely overcorrects again in this case). Comparison of TD-CBT to the wait-list (WL) is inconsistent for both depression and anxiety across all three analyses. While the original analysis rejects the null hypotheses with $p < 0.001$ and finds substantial improvement in the TD-CBT group (1.23 and 1.24 for depression and anxiety respectively), the RoBMA-regression does find strong and moderate evidence for the presence of improvements in the TD-CBT group, however with the publication bias-adjusted model-averaged effect sizes of approximately 1/3 (0.38 and 0.37) of the original effect size estimates. Furthermore, RoBMA_{PSMA} subgroup analyses show weak evidence in favor of no improvements in the TD-CBT group and further shrink the effect sizes to tenths of the original degree (0.12 and 0.15).

Differences between the original subgroup analyses and the sensitivity analyses are indicative of the degree of publication bias. The additional differences between the RoBMA-regression sensitivity analyses and RoBMA_{PSMA} subgroup analyses suggest that the publication bias and heterogeneity differ across the comparison groups. In fact, while RoBMA-regression found extreme evidence of publication bias in both depression and anxiety, the RoBMA_{PSMA} subgroup analyses did not find evidence of publication bias in comparison of TD-CBT to DS-CBT on depression and only weak evidence in comparison of TD-CBT to other comparison groups on anxiety.

4 Conclusion

The robustness analyses question the added benefit of TD-CBT on treating depression and anxiety. Contrary to the original analyses, which indicated significant

overall effects, the robustness analyses revealed weak evidence for the absence of these effects when accounting for publication bias. While we cannot rule a partial influence of small-study effects on the publication bias adjustment in anxiety, the highly reduced overall effect sizes and strong evidence of publication bias highlight the need to interpret the original findings cautiously. Furthermore, substantial between-study heterogeneity suggests that reported effects differ across comparison groups. While the original results indicated that TD-CBT is more effective than treatment as usual, other control groups (for depression), and wait-list, the results of the publication bias adjusted meta-regression and subgroup analyses show inconsistent conclusions and notably reduced effect size estimates.

References

- Bartoš, F., Maier, M., Stanley, T. and Wagenmakers, E.-J.: 2024, Robust Bayesian meta-regression: Model-averaged moderation analysis in the presence of publication bias.
- Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H. and Stanley, T. D.: 2022, Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods, *Research Synthesis Methods* **14**(1), 99–116.
- Bartoš, F. and Maier, M.: 2020, RoBMA: An R package for robust Bayesian meta-analyses. R package version 3.2.0.
- Bom, P. R. and Rachinger, H.: 2020, A generalized-weights solution to sample overlap in meta-analysis, *Research Synthesis Methods* **11**(6), 812–832.
- Brodeur, A., Dreber, A., Hoces de la Guardia, F. and Miguel, E.: 2024, Reproduction and replication at scale, *Nature Human Behaviour* **8**(1), 2–3.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M. and Hilgard, J.: 2019, Correcting for bias in psychology: A comparison of meta-analytic methods, *Advances in Methods and Practices in Psychological Science* **2**(2), 115–144.
- Hinne, M., Gronau, Q. F., van den Bergh, D. and Wagenmakers, E.-J.: 2020, A conceptual introduction to Bayesian model averaging, *Advances in Methods and Practices in Psychological Science* **3**(2), 200–215.
- Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T.: 1999, Bayesian model averaging: a tutorial, *Statistical Science* **14**(4), 382–401.
- Hong, S. and Reed, W. R.: 2020, Using monte carlo experiments to select meta-analytic estimators, *Research Synthesis Methods* **12**(2), 192–215.
- JASP Team: 2024, JASP (Version 0.19.2)[Computer software].
URL: <https://jasp-stats.org/>
- Jeffreys, H.: 1939, *Theory of probability*, 1st edition edn, Oxford University Press, Oxford, UK.
- Kvarven, A., Strømmland, E. and Johannesson, M.: 2020, Comparing meta-analyses and preregistered multiple-laboratory replication projects, *Nature Human Behaviour* **4**(4), 423–434.
- Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H. and Olkin, I.: 2006, The case of the misleading funnel plot, *BMJ* **333**(7568), 597–600.
- Lee, M. D. and Wagenmakers, E.-J.: 2013, *Bayesian cognitive modeling: A practical course*, Cambridge University Press, New York.
- Maier, M., Bartoš, F. and Wagenmakers, E.-J.: 2023, Robust Bayesian meta-analysis: Addressing publication bias with model-averaging, *Psychological Methods* **28**(1), 107–122.

- Schaeuffele, C., Meine, L. E., Schulz, A., Weber, M. C., Moser, A., Paersch, C., Recher, D., Boettcher, J., Renneberg, B., Flückiger, C. et al.: 2024, A systematic review and meta-analysis of transdiagnostic cognitive behavioural therapies for emotional disorders, *Nature Human Behaviour* **8**(3), 493–509.
- Stanley, T. D. and Doucouliagos, H.: 2014, Meta-regression approximations to reduce publication selection bias, *Research Synthesis Methods* **5**(1), 60–78.
- Vevea, J. L. and Hedges, L. V.: 1995, A general linear model for estimating effect size in the presence of publication bias, *Psychometrika* **60**(3), 419–435.
- Viechtbauer, W.: 2010, Conducting meta-analyses in R with the metafor package, *Journal of Statistical Software* **36**(3), 1–48.

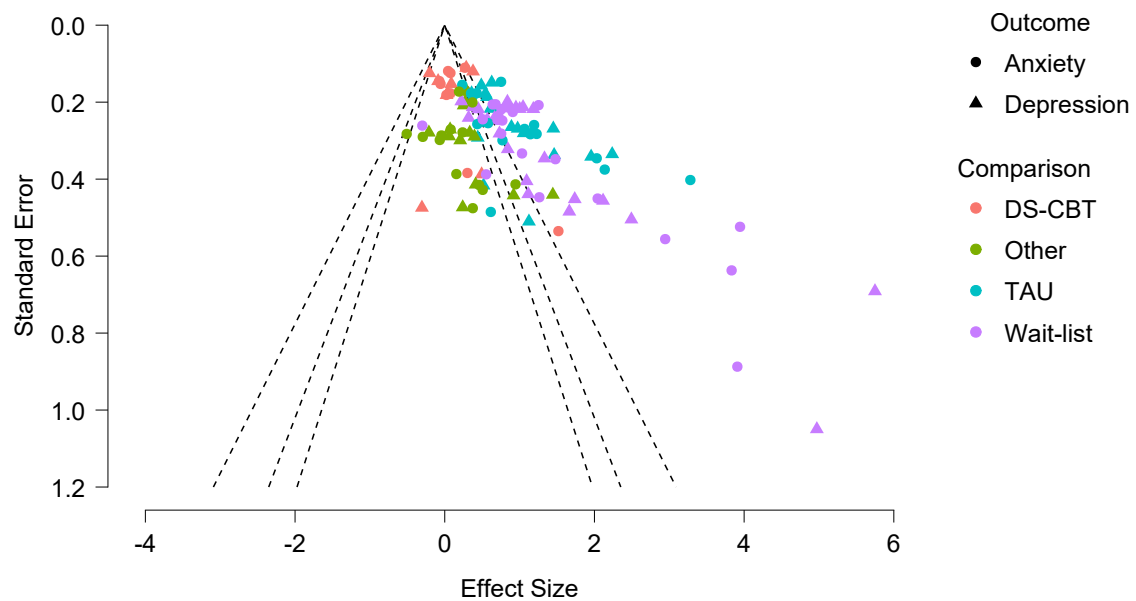


Figure 1: Funnel Plot of Effect Size Estimates of the Transdiagnostic Cognitive Behavioural Therapies on Depression (Triangles) and Anxiety (Circles) from [Schaeuffele et al. \(2024\)](#) Shows Extensive Publication Bias. Funnels with 90%, 95%, and 99% Prediction Intervals Are Centered at 0. Figure from JASP ([JASP Team 2024](#)).

5 Figures

6 Tables

Table 1: Comparison of the Summary Information Data Extraction from Selected Studies by [Schaeuffele et al. \(2024\)](#) and Authors.

Study	Measure	Original			Recoded		
		m _{ctrl}	m _{td}	n _{ctrl}	m _{ctrl}	m _{td}	n _{ctrl}
		sd _{ctrl}	sd _{td}	n _{td}	sd _{ctrl}	sd _{td}	n _{td}
Depression							
Norton & Hope (2005)	DASS-D	23.3	5.44	9 [†]	23.3	5.44	10
		3.31	3.49	6 [†]	3.31	3.49	9
Zemestani et al. (2017)	BDI-II	25.8	11.60	23	25.8	11.60	23
		2.80	1.90	20	2.80	1.90	20
Anxiety							
Nazari et al. (2020b)	HADS-A	12.06 [†]	7.25 [†]	28 [†]	11.69	8.09	23
		1.26 [†]	1.60 [†]	30 [†]	1.65	1.82	28
Nasiri et al. (2020)	BAI	34.00	9.54	15	34.00	9.54	15
		6.50	5.82	13	6.50	5.82	13
Norton & Hope (2005)	DASS-A	17.00	3.77	9 [†]	17.00	3.77	10
		3.11	3.30	6 [†]	3.11	3.30	9
Zemestani et al. (2017)	BAI	20.80	11.10	23	20.80	11.10	23
		3.10	1.20	20	3.10	1.20	20

Notes: Results of the data extraction by [Schaeuffele et al. \(2024\)](#) (Original) and us (Recoded) for means (m), standard deviations (sd) and sample sizes (n) in the control (ctrl) and the treatment (trt) groups. All values are post-treatment. Discrepancies between the original and recoded values are marked with a [†] at the original coding. Therapy setting (group vs. individual) and the type of control group are not displayed as no discrepancies were found.

Table 2: Planned Publication Bias Adjustment Robustness Analyses of the Overall Effect of Transdiagnostic Cognitive Behavioural Therapies on Depression and Anxiety.

	Depression		Anxiety	
	Estimate	Test	Estimate	Test
Original	0.74 (0.57, 0.92)	$p < 0.001$ reject H_0	0.77 (0.56–0.97)	$p < 0.001$ reject H_0
Robustness 1	0.05	$BF_{10} = 0.34$	0.06	$BF_{10} = 0.42$
RoBMA _{PSMA}	(−0.13, 0.50)	weak for H_0	(−0.19, 0.57)	weak for H_0
Robustness 2	0.06	$BF_{10} = 0.36$	0.08	$BF_{10} = 0.45$
+ Down-weighting	(−0.16, 0.55)	weak for H_0	(−0.17, 0.63)	weak for H_0
Robustness 3	0.07	$BF_{10} = 0.38$	0.12	$BF_{10} = 0.61$
– Outliers	(−0.03, 0.57)	weak for H_0	(−0.03, 0.62)	weak for H_0

Notes: Original refers to the results reported by [Schaeuffele et al. \(2024\)](#). Robustness 1 corresponds to the first robustness analysis using RoBMA_{PSMA}. Robustness 2 extends the first robustness analysis by accounting for sample overlap in the comparison group. Robustness 3 extends the second robustness analysis by removing potential outliers (effect sizes larger than $d = 3$). The Estimate column shows overall effect size estimates with 95% confidence/credible intervals in brackets. The Tests columns show the test of the overall effect size estimate with the appropriate interpretation.

Table 3: Sensitivity Analysis of the Publication Bias Adjustment of the Overall Effect of Transdiagnostic Cognitive Behavioural Therapies on Depression and Anxiety.

	Depression		Anxiety	
	Estimate	Test	Estimate	Test
RoBMA with Selection Models	0.10 (-0.22, 0.70)	$\text{BF}_{10} = 0.59$ weak for H_0	0.52 (0.00, 0.87)	$\text{BF}_{10} = 5.4$ moderate for H_1
RoBMA with PET and PEESE	0.05 (-0.13, 0.48)	$\text{BF}_{10} = 0.32$ moderate for H_0	0.06 (-0.19, 0.53)	$\text{BF}_{10} = 0.40$ weak for H_0

Notes: RoBMA with Selection Models corresponds to RoBMA_{PSMA} omitting PET and PEESE style models from the publication bias adjustment part. RoBMA with PET and PEESE corresponds to RoBMA_{PSMA} omitting selection models from the publication bias adjustment part. The Estimate column shows overall effect size estimates with 95% credible intervals in brackets. The Tests columns show the test of the overall effect size estimate with the appropriate interpretation.

Table 4: Sensitivity Analyses of the by Comparison-Group Subgroup Analyses of Transdiagnostic Cognitive Behavioural Therapies on Depression and Anxiety.

	Depression		Anxiety	
	Estimate	Test	Estimate	Test
vs. DS-CBT				
Original	0.09 (−0.07, 0.25)	$p = 0.269$ not reject H_0	0.09 (−0.01, 0.20)	$p = 0.091$ not reject H_0
RoBMA-regression	−0.24 (−0.52, 0.14)	$BF_{10} = 1.4$ weak for H_1^*	−0.19 (−0.55, 0.18)	$BF_{10} = 0.74$ weak for H_0
RoBMA _{PSMA}	0.01 (0.00, 0.15)	$BF_{10} = 0.15$ moderate for H_0	0.00 (−0.01, 0.10)	$BF_{10} = 0.11$ moderate for H_0
vs. TAU				
Original	0.90 (0.66, 1.14)	$p < 0.001$ reject H_0	0.98 (0.63, 1.33)	$p < 0.001$ reject H_0
RoBMA-regression	0.28 (0.03, 0.67)	$BF_{10} = 3.1$ moderate for H_1	0.32 (−0.08, 0.69)	$BF_{10} = 3.0$ moderate for H_1
RoBMA _{PSMA}	0.48 (0.00, 1.00)	$BF_{10} = 2.8$ weak for H_1	0.33 (−0.35, 1.06)	$BF_{10} = 1.5$ weak for H_1
vs. Other				
Original	0.27 (0.13, 0.42)	$p < 0.001$ reject H_0	0.14 (−0.04, 0.31)	$p = 0.128$ not reject H_0
RoBMA-regression	−0.28 (−0.57, 0.16)	$BF_{10} = 1.9$ weak for H_1^*	−0.46 (−0.92, −0.05)	$BF_{10} = 5.1$ moderate for H_1^*
RoBMA _{PSMA}	0.07 (−0.05, 0.37)	$BF_{10} = 0.57$ weak for H_0	0.02 (−0.03, 0.20)	$BF_{10} = 0.21$ moderate for H_0
vs. WL				
Original	1.23 (0.80, 1.66)	$p < 0.001$ reject H_0	1.24 (0.82, 1.67)	$p < 0.001$ reject H_0
RoBMA-regression	0.38 (0.12, 0.78)	$BF_{10} = 14.3$ strong for H_1	0.37 (−0.03, 0.76)	$BF_{10} = 4.7$ moderate for H_1
RoBMA _{PSMA}	0.12 (−0.56, 0.83)	$BF_{10} = 0.86$ weak for H_0	0.15 (−0.56, 0.90)	$BF_{10} = 0.96$ weak for H_0

Notes: Original refers to the subgroup analysis results reported by [Schaeuffele et al. \(2024\)](#) in Table 2. RoBMA-regression corresponds to marginal mean estimates from robust Bayesian meta-regression. RoBMA_{PSMA} corresponds to subgroup analyses performed by RoBMA_{PSMA}. The Estimate column shows effect size estimates with 95% confidence/credible intervals in brackets. The Tests columns show the test of the effect size estimate with the appropriate interpretation.

Table 5: Replication Package Contents and Reproducibility

Replication Package Item	Fully	Partial	No
Raw data provided	✓		
Analysis data provided	✓		
Cleaning code provided	✓		
Analysis code provided	✓		
Reproducible from raw data	✓		
Reproducible from analysis data	✓		

Notes: This table summarizes the replication package contents contained in [Schaeuffele et al. \(2024\)](#).