Verification Report of Zhao et al. (2023) 'Effect of Acceptance and Commitment Therapy for depressive disorders: a meta-analysis'

Ian Hussey

Zhao et al.'s (2023) recent meta-analysis of ACT vs. control interventions for depressive disorders reported an overall meta-analytic effect size of Hedges g = -1.05. However, errors appear to have been made in the extraction of effect sizes. Re-extraction of the effect sizes from the original articles suggests that 3 of 8 recalculated effect sizes were erroneous: Standard Errors (SE) were confused for Standard Deviations (SD) in two cases, and incorrect sample sizes were extracted in one case. When the first meta-analysis they report was re-run following Zhao et al.'s (2023) analytic strategy, the resulting meta-analytic effect size of Hedges' g = 0.68 was 35% smaller than that reported in the original meta-analysis. The results of other subgroup meta-analyses reported by Zhao et al. (2023) are likely also to be affected. Other aspects of Zhao et al. (2023) may also require scrutiny for the presence of errors. Given the detected errors in the results, the conclusions of Zhao et al. (2023) are undermined and may require correction. All data and code available at osf.io/jwcmu.

Zhao et al. (2023) present what they describe as "the most up to date meta-analysis examining the effect of ACT [Acceptance and Commitment Therapy] for patients with depressive disorders." (p. 10). Meta-analyses are often argued to be near the top of the hierarchy of evidence and are influential in determining whether a given therapy is considered evidence-based, and how its efficacy compares to other therapies (Tolin et al., 2015).

Elsewhere, however, numerous authors have highlighted the fact that meta-analyses frequency contain errors and have poor reproducibility (Gøtzsche et al., 2007; Lakens et al., 2017; López-Nicolás et al., 2022). For example, among psychological meta-analyses, 45% of effect sizes (224 of 500) could not be reproduced from the original articles, leading to discrepancies in the meta-analysis results in 39% of cases (13 of 33: Maassen et al., 2020). These errors are not random but often take common and preventable forms (Kadlec et al., 2023). For example, 45% of highly-cited meta-analyses in strength and conditioning research miscalculate at least one Standardized Mean Difference effect sizes (e.g., Cohen's d and Hedges' g) from Standard Errors (SE) instead of Standard Deviations (SD), thereby greatly inflating the effect size estimate and distorting the meta-analytic effect size.

Recently, the concept of the Verification Report has been introduced (Chambers, 2020): articles whose primary goal is to verify the analyses reported in previous research so as to assess the accuracy, robustness, and credibility of that work. Given the importance of meta-analyses for assessing the efficacy of psychotherapy, and at the same time the strong a priori probability that such meta-analyses contain errors, I therefore sought to verify some of the key analyses reported by Zhao et al. (2023).

Table 1. Influence metrics for Zhao et al.'s (2023) effect sizes

Article	rstudent	DFFITS	Cook's distance	Cov ratio	$\Delta\tau^2$	ΔQE	hat	weight	DFBETAS	Outlier
Wendy T. M. Pots 2016	0.20	0.06	0.00	1.34	1.04	140.42	0.10	9.76	0.06	
Situ Yuyi 2022	-0.30	-0.10	0.01	1.20	0.94	144.70	0.09	9.25	-0.10	
Shima Tamannaei Far 2017	1.51	0.44	0.19	1.04	0.81	133.49	0.08	7.97	0.44	
Ren Zhihong 2012	-0.53	-0.18	0.04	1.31	1.02	138.61	0.10	9.75	-0.18	
Mehdi Zemestani 2020	-2.11	-0.64	0.37	0.97	0.74	120.95	0.09	8.70	-0.64	
Louise Hayes 2011	-0.50	-0.15	0.02	1.14	0.89	144.26	0.08	8.40	-0.15	
Lappalainen 2015	0.13	0.04	0.00	1.17	0.91	144.45	0.09	8.92	0.04	
Ernst T. Bohlmeijer 2011	0.16	0.05	0.00	1.24	0.97	143.33	0.10	9.50	0.05	
Chunxiao Zhao 2022	-0.53	-0.18	0.04	1.31	1.02	138.61	0.10	9.75	-0.18	
Chen Juan 2021	-1.09	-0.34	0.12	1.12	0.86	136.07	0.09	9.09	-0.34	
A-Tjak 2018	4.74	1.52	1.03	0.47	0.32	56.20	0.09	8.91	1.55	*

Note: For explanations of each influence metric see the metafor documentation:
https://wviechtb.github.io/metafor/reference/influence.rma.uni.html

Figure 1. Reproduction of Zhao et al.'s (2023) meta-analysis sensitivity analysis from the summary statistics reported in their forest plot (Figure 4).
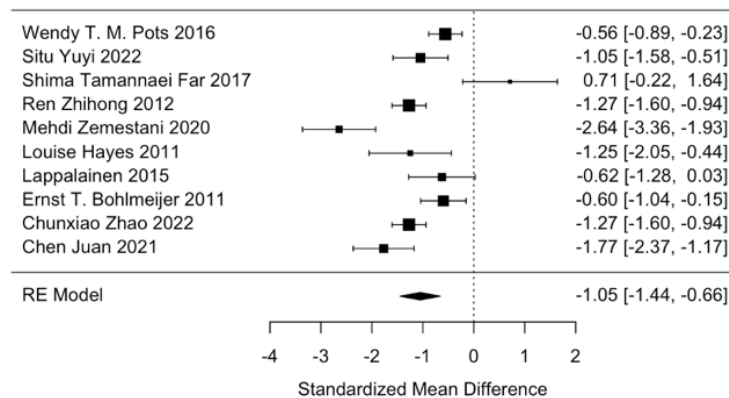
Attempt to reproduce the first meta-analysis

Zhao et al. (2023) presented the results of multiple meta-analyses, associated sensitivity analyses, and subgroup analyses. Here, I attempt to reproduce and verify the result of the first meta-analysis reported and its associated forest plot (Figure 4, p. 6), which was reported to include the full set of results from studies found to match the inclusion and exclusion criteria. The rationale for scrutinizing this set of results was that any issues detected there would likely have downstream implications for the other meta-analyses conducted on subgroups.

Under the subheading "Effect of ACT on depression", Zhao et al. (2023, p. 6) reported that *"Eleven studies reported the effect of ACT on depression levels in 887 patients with depressive disorders. There was statistically significant heterogeneity among the studies (p < 0.00001; I²= 93%), therefore, the random effects model was used to conduct meta-analysis. There was significant treatment effect from ACT in comparison to other treatments, as shown in the forest plot in Fig. 4."* However, Figure 4 does not actually present the results from this meta-analysis but rather, as its title states, "Effect of ACT on depression in patients with depressive disorders after sensitivity analysis". Zhao et al. (2023) excluded A-Tjak et al. (2018) from the meta analysis, although they did not state their rationale for doing so.

In order to attempt to reproduce the results of the meta-analysis, I first extracted the summary statistics (M, SD, and N for both intervention and

control groups) from Figure 4. Standardized Mean Difference effect sizes were then recalculated using the metafor package's (Viechtbauer, 2010) escalc function. A random effects meta-analysis model was then fit using metafor's rma function using the DerSimonian-Laird estimator, on the basis that Zhao et al. (2023) fit their analyses in Revman (The Cochrane Collaboration, n.d.) which uses that estimator by default. All data and code for the verification can be found on the Open Science Framework (osf.io/jwcmu).

Although Zhao et al. (2023) did not state their specific rationale for excluding A-Tjak et al. (2018) from their meta-analysis, it was nonetheless possible to check whether I arrived at a similar conclusion. Using metafor's influence function to detect effect sizes with undue influence on the results (i.e., outliers) flagged A-Tjak et al. (2018) as the only likely outlier (see Table 1)[1]. Broadly speaking, this analytic choice (although its specific method in the original publication remains indeterminable) could therefore be reproduced.

Next, I fit a meta-analytic model to the remaining 10 studies to try to reproduce the meta-analysis sensitivity analysis results reported in Zhao et al. (2023) Figure 4. For comparison, the original results reported by Zhao et al. (2023, Figure 4) were meta-effect size[2]: Hedges' g = -1.05, 95% [-1.44, -0.66], $p < .00001$. Heterogeneity: $\tau^2 = 0.31$, $I^2 = 84\%$. The results of the verification analysis were almost identical: Hedges' g = -1.05, 95% [-1.44, -0.66], $p < .0001$. Heterogeneity: $\tau^2 = 0.32$, $I^2 = 84\%$. That is, there was a .01 discrepancy between the original τ^2 and the recalculated result. These results were therefore substantively verified with discrepancies within what could be accounted for by rounding differences.

Attempt to reproduce the summary statistics and effect sizes employed in the meta-analysis

Although the sensitivity analysis results reported in Figure 4 were substantively reproduced from the summary statistics reported in that figure, the validity of those results are reliant on the summary statistics being accurately extracted from the original studies. Recent work has shown that this is frequently not the case (e.g., Maassen et al., 2020). I therefore attempted to extract the Mean, SD, and N for the ACT and control conditions at the post-intervention time point for each of the original studies.

Two articles with DOIs were only available in Chinese for a fee (Chen et al., 2021; Situ, 2022)c and were not accessed. One thesis (Ren, 2012) did not have a DOI listed and could not be found via internet searches. It appears to be a thesis. Notably, Zhao et al. (2023) stated their search strategy as "Studies published in both English and Chinese in peer reviewed journals were searched. We also examined reference lists of appropriate articles in order to identify additional relevant studies." (p. 2), however their PRISMA flow chart (p. 4) does not state that they included any unpublished studies from other sources. The provenance of Ren (2012) is therefore unclear. In summary, the reproducibility of 3 of the 11 studies' effect sizes was indeterminable.

In 3 of the remaining 8 articles (37.5%), the summary statistics (M, SD, or N) reported in Zhao et al. (2023, Figure 4) did not match those re-extracted from the original studies. In two cases (A-Tjak et al., 2018; Hayes et al., 2011), the original articles reported Standard Errors (SE) which were incorrectly employed as Standard Deviations (SD) by Zhao et al. (2023) as if they were Standard Deviations. This is a common and unfortunate issue in meta analyses (Kadlec et al., 2023), and is problematic as SEs are much smaller than SDs which inflates the calculated Standardized Mean Difference effect sizes. In the remaining case (Zemestani & Mozaffari, 2020), the sample sizes do not match those reported in the original publication. This also distorts SMD effect sizes, although in this case by a much smaller degree. In summary, only 5 of the 11 sets of summary statistics could be verified against the numbers reported in the original articles. Both Zhao et al.'s (2023) summary statistics and the re-extracted summary statistics are reported in Table 2.

---

[1] Note that in all plots and tables I use the same labelling as Zhao et al. (2023). Note that there is imperfect correspondence between Zhao's labels and the authors of the original studies. The references (including DOIs) for the studies they refer to are available in the Supplementary Materials.

[2] Note that Zhao et al. (2023) were not explicit about whether they applied Hedge's correction to their Standardized Mean Difference effect sizes, although they do mention it in Figure 3. Because of this and the fact that it results in an equivalent effect size in my reanalysis, I employ it throughout.

Table 2. Comparison of summary statistics reported in Zhao et al.'s (2023) forest plot (Figure 4) vs. (b) re-extracted from the original articles.

Article	Source	ACT condition			Control condition			Comparison
		M	SD	N	M	SD	N	
Wendy T. M. Pots 2016	Zhao et al.	14.68	8.05	71	19.34	8.55	78	Consistent
	Re-extracted	14.68	8.05	71	19.34	8.55	78	
Situ Yuyi 2022	Zhao et al.	12.51	3.26	30	15.52	2.35	30	Indeterminable
	Re-extracted	-	-	-	-	-	-	
Shima Tamannaei Far 2017	Zhao et al.	28.20	16.28	10	18.54	7.65	9	Consistent
	Re-extracted	28.20	16.28	10	18.54	7.65	9	
Ren Zhihong 2012	Zhao et al.	13.61	9.48	92	26.05	10.06	76	Indeterminable
	Re-extracted	-	-	-	-	-	-	
Mehdi Zemestani 2020	Zhao et al.	20.70	3.40	26	32.57	5.15	30	Inconsistent*
	Re-extracted	20.70	3.40	23	32.57	5.15	29	
Louise Hayes 2011	Zhao et al.	66.05	3.24	19	70.68	4.20	11	Inconsistent**
	Re-extracted	66.05	14.12	19	70.68	13.93	11	
Lappalainen 2015	Zhao et al.	13.34	6.75	18	17.85	7.34	20	Consistent
	Re-extracted	13.34	6.75	18	17.85	7.34	20	
Ernst T. Bohlmeijer 2011	Zhao et al.	15.94	10.37	39	22.07	9.99	42	Consistent
	Re-extracted	15.94	10.37	39	22.07	9.99	42	
Chunxiao Zhao 2022	Zhao et al.	13.61	9.48	92	26.05	10.06	76	Consistent
	Re-extracted	13.61	9.48	92	26.05	10.06	76	
Chen Juan 2021	Zhao et al.	49.36	2.18	30	54.36	3.29	30	Indeterminable
	Re-extracted	-	-	-	-	-	-	
A-Tjak 2018	Zhao et al.	11.08	1.36	33	7.91	1.51	25	Inconsistent**
	Re-extracted	11.08	7.81	33	7.91	7.55	25	

* Incorrectly extracted result
** Erroneously treated a Standard Error (SE) as a Standard Deviation (SD)

Figure 2. Comparison of effect size estimates calculated from summary statistics (a) reported in Zhao et al.'s (2023) forest plot (Figure 4) vs. (b) re-extracted from the original articles.
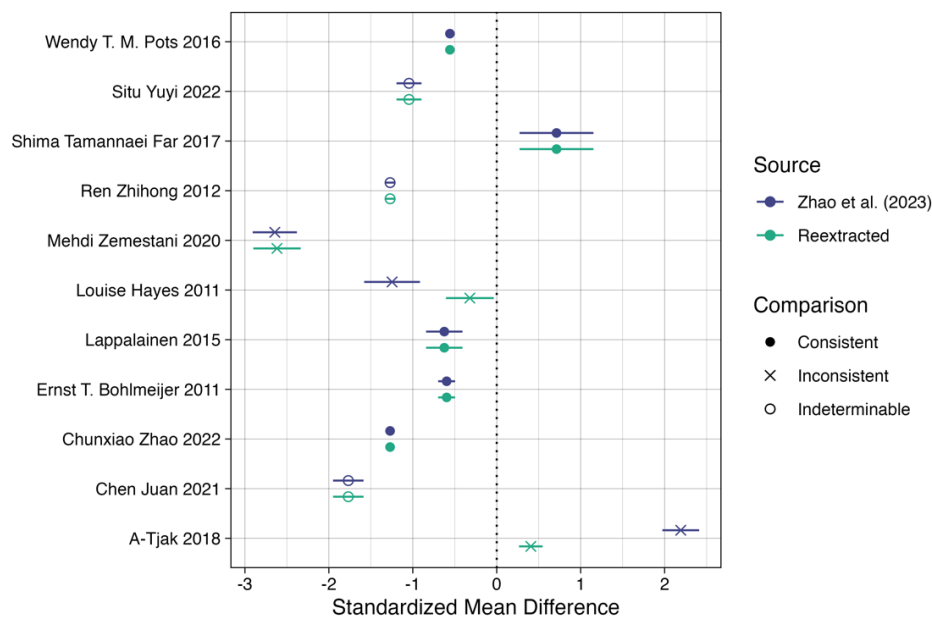
Table 3. Influence metrics for the recalculated effect sizes

Article	rstudent	DFFITS	Cook's distance	Cov ratio	$\Delta\tau^2$	ΔQE	hat	weight	DFBETAS	Outlier
Wendy T. M. Pots 2016	0.39	0.12	0.02	1.28	0.52	77.78	0.10	10.20	0.12	
Situ Yuyi 2022	-0.30	-0.10	0.01	1.20	0.49	82.46	0.09	9.25	-0.10	
Shima Tamannaei Far 2017	2.05	0.56	0.29	0.96	0.39	71.01	0.07	7.19	0.57	
Ren Zhihong 2012	-0.63	-0.22	0.05	1.26	0.51	76.78	0.10	10.18	-0.22	
Mehdi Zemestani 2020	-2.69	-0.76	0.47	0.86	0.33	61.35	0.08	8.18	-0.77	*
Louise Hayes 2011	0.69	0.21	0.04	1.12	0.45	80.43	0.08	8.16	0.20	
Lappalainen 2015	0.29	0.08	0.01	1.16	0.47	82.10	0.09	8.67	0.08	
Ernst T. Bohlmeijer 2011	0.34	0.10	0.01	1.22	0.50	80.88	0.10	9.71	0.10	
Chunxiao Zhao 2022	-0.63	-0.22	0.05	1.26	0.51	76.78	0.10	10.18	-0.22	
Chen Juan 2021	-1.36	-0.42	0.17	1.06	0.42	74.05	0.09	8.96	-0.42	
A-Tjak 2018	2.08	0.69	0.37	0.85	0.32	57.34	0.09	9.33	0.69	

Note: For explanations of each influence metric see the metafor documentation:
https://wviechtb.github.io/metafor/reference/influence.rma.uni.html

Figure 3. Forest plot for corrected meta-analysis sensitivity analysis.
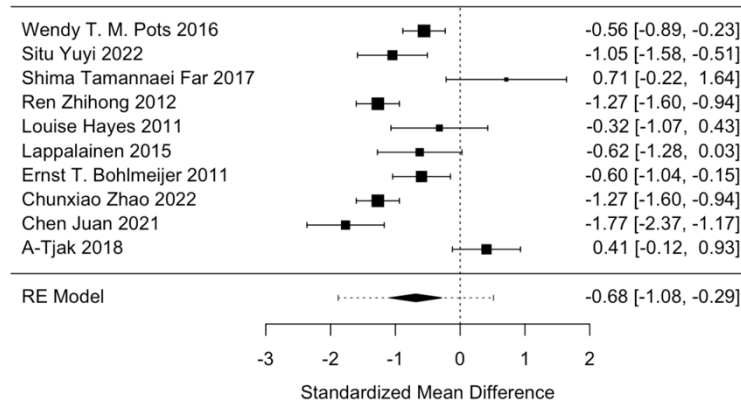
Next, metafor's escalc function was again used to calculate SMD effect sizes (Hedges' g) and its variances. Recalculated Hedge's g values differed from Zhao et al.'s (2023) by up to $g = 1.78$. Put differently, the recalculated values were as little as 18.7% of the those calculated by Zhao et al. (2023). This is notable for two reasons. First, the original article this effect size was calculated from (A-Tjak et al., 2018) also reported an SMD effect size directly, but the discrepancy between A-Tjak et al.'s effect size ($d = 0.25$) and Zhao et al.'s (2023) extracted effect size ($g = 2.19$) apparently did not prompt the meta-analysts to re-examine whether it had been calculated correctly. Second, this effect size was excluded as an outlier by Zhao et al. (2023), but this apparently did not prompt the meta-analysts to re-examine whether it had

been calculated correctly either. Both sets of effect sizes are illustrated in Figure 2.

Corrected meta-analysis

Following the same workflow as Zhao et al. (2023), a random effects meta-analysis was fit to all 11 effect sizes calculated from the summary statistics. The re-extracted summary statistics were used where possible. Where the original article could not be obtained, Zhao et al.'s (2023) summary statistics were used on the liberal assumption that they are correct.

When comparing the meta-analysis models fit to all 11 effect sizes, heterogeneity was lower in the reextracted effect sizes ($\tau^2 = 0.44$, $I^2 = 87.9\%$, $H^2 = 8.28$) than Zhao et al.'s (2023) reported effect sizes ($\tau^2 = 0.85$, $I^2 = 93.1\%$, $H^2 = 14.51$). This is to be expected when some of Zhao et al.'s (2023) effect sizes were erroneously inflated.

Following Zhao et al.'s (2023) analytic strategy, test of undue influence/outliers were calculated. This was done using metafor's influence function. The effect size from Zemestani & Mozaffari (2020) was flagged as an outlier, and therefore excluded from the sensitivity analysis.

Note that there are also other reasons to exclude the results of Zemestani & Mozaffari (2020). As documented in a comment I wrote on PubPeer in February 2024 (https://pubpeer.com/publications/0E13E34679 B18385D6C4C29143A9CD), several results reported in that article appear to be mathematically impossible as they fail StatCheck (Nuijten et al., 2015; Nuijten & Polanin, 2020), GRIM (Brown & Heathers, 2017), or GRIMMER tests (Anaya, 2016); the reported SDs of the BDI-II scores (3.4 and 5.15) are implausibly small for BDI-II, which should be around 8 to 12 (Wang & Gorenstein, 2013), and as a result the SMD effect size is implausibly extreme (Hedges' g = -2.64). I also previously contacted the authors of Zemestandi & Mozaffari (2020) by email in October 2023 and January 2024 to raise my concerns privately and seek clarification from them. Unfortunately, I received no response to my emails and no replies have been posted to PubPeer to address these concerns in the seven months between writing the PubPeer comment (February 2024) and writing of this Verification Report (September 2024). As such, the Cochrane handbook recommends that such results be excluded on the basis that they are currently 'awaiting assessment' (https://training.cochrane.org/handbook/current /chapter-05#section-5-5-10).

The remaining 10 effect size estimates were then fit to a meta-analytic model, as before. The results demonstrated a meta-effect size of Hedges' g = -0.68, 95% [-1.08, -0.29], p = .0007. Heterogeneity was τ^2 = 0.33, I^2 = 85%. As such, the corrected meta-analysis still found a statistically significant standardized effect size, however the magnitude of this effect was -35.0% smaller than that reported by Zhao et al. (2023) after correcting data extraction errors and reapplying outlier exclusions. The impact of these errors can also be expressed via the difference in the Number Needed to Treat (NNT: Cook & Sackett, 1995) between the original and corrected results: the original results suggest that an average of 1.84 people need to be treated for one additional person to benefit, whereas the corrected meta-analysis suggests that 2.69 people need to be treated for one to receive benefit. This 46% increase in NNT greatly changes the cost-benefit analysis of ACT for depression in terms of the resources per person who benefits from the intervention.

Discussion

The results of Zhao et al.'s (2023) meta-analyses appear to be based on erroneously extracted and miscalculated summary statistics that follow common patterns of error in the meta-analysis literature such as confusing Standard Errors (SE) for Standard Deviations (SD; see Kadlec et al., 2023). When the first meta-analysis they report was re-run following Zhao et al.'s (2023) analytic strategy, the resulting meta-analytic effect size of Hedges' g = 0.68 was 35% smaller than that reported in the original meta-analysis (g = -1.05).

These types of errors are both prevalent and preventable. Zhao et al. (2023) appear to have not detected them despite multiple red flags being raised: the original articles reported effect sizes that were very different from Zhao et al.'s (2023) estimates, and at least one erroneous effect size was detected as an outlier. However, neither red flag apparently prompted the authors to revisit their effect size extractions. Other warning signs were detected here, such as concerns about the veracity of the results reported for one article included in Zhao et al.'s (2023) meta-analysis. Future work may be aided by recent developments in the detection and exclusion of such problematic data, such as the INSPECT-SR tool that is in development for Cochrane reviews (Wilkinson et al., 2023).

Limitations

As discussed at the start of this document, no attempt was made to reproduce or attempt to understand the impact of the detected extraction errors on the other (meta-)analyses reported in Zhao et al. (2023), such as their subgroup analyses. It is likely that those subgroup analyses reported after Figure 4 are also be affected by the issues detected here. It is also possible that other issues exist, given that errors were made at the relatively simple data extraction phase.

Equally, this verification does not attempt to critique things that were not done by Zhao et al. (2023). For example, the authors do not apply any method of bias correction for p-hacking or publication bias, and instead take the reported results at face value. Applying such tests typically

shrinks the estimate of the true effect (Carter et al., 2019).

Conclusion

Given the detected errors in the results, the conclusions of Zhao et al. (2023) are undermined and may require correction.

Author notes

Ian Hussey (corresponding author: ian.hussey@unibe.ch), University of Bern, Switzerland. ORCID 0000-0001-8906-7559.

References

Anaya, J. (2016). *The GRIMMER test: A method for testing the validity of reported measures of variability* (e2400v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.2400v1

A-Tjak, J. G. L., Morina, N., Topper, M., & Emmelkamp, P. M. G. (2018). A Randomized Controlled Trial in Routine Clinical Practice Comparing Acceptance and Commitment Therapy with Cognitive Behavioral Therapy for the Treatment of Major Depressive Disorder. *Psychotherapy and Psychosomatics.* https://doi.org/10.1159/000486807

Brown, N. J. L., & Heathers, J. A. J. (2017). The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology. *Social Psychological and Personality Science, 8*(4), 363–369. https://doi.org/10.1177/1948550616673876

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & 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. https://doi.org/10.1177/2515245919847196

Chambers, C. D. (2020). Verification Reports: A new article type at Cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 129,* A1–A3. https://doi.org/10.1016/j.cortex.2020.04.020

Chen, J., Gao, Y., Yuan, X. J., & Wu, J. (2021). Analysis of the effect of acceptance and commitment therapy in the psychological nursing application of patients with depression. *Psychology Monthly, 16*(14), 49–50. https://doi.org/10.19738/j.cnki.psy.2021.14.020

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: A clinically useful measure of treatment effect. *BMJ: British Medical Journal, 310*(6977), 452–454.

Gøtzsche, P. C., Hróbjartsson, A., Marić, K., & Tendal, B. (2007). Data Extraction Errors in Meta-analyses That Use Standardized Mean Differences. *JAMA, 298*(4), 430–437. https://doi.org/10.1001/jama.298.4.430

Hayes, L., Boyd, C. P., & Sewell, J. (2011). Acceptance and Commitment Therapy for the Treatment of Adolescent Depression: A Pilot Study in a Psychiatric Outpatient Setting. *Mindfulness, 2*(2), 86–94. https://doi.org/10.1007/s12671-011-0046-5

Kadlec, D., Sainani, K. L., & Nimphius, S. (2023). With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. *Sports Medicine, 53*(2), 313–325. https://doi.org/10.1007/s40279-022-01766-0

Lakens, D., Page-Gould, E., van Assen, M. A. L. M., Spellman, B., Schönbrodt, F. D., Hasselman, F., Corker, K. S., Grange, J., Sharples, A., Cavender, C., Augusteijn, H., Augusteijn, H., Gerger, H., Locher, C., Miller, I. D., Anvari, F., & Scheel, A. M. (2017). *Examining the Reproducibility of Meta-Analyses in Psychology: A Preliminary Report* [Preprint]. BITSS. https://doi.org/10.31222/osf.io/xfbjf

López-Nicolás, R., López-López, J. A., Rubio-Aparicio, M., & Sánchez-Meca, J. (2022). A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000–2020). *Behavior Research Methods, 54*(1), 334–349. https://doi.org/10.3758/s13428-021-01644-z

Maassen, E., Assen, M. A. L. M. van, Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLOS ONE, 15*(5), e0233107. https://doi.org/10.1371/journal.pone.0233107

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods.* https://doi.org/10.3758/s13428-015-0664-2

Nuijten, M. B., & Polanin, J. R. (2020). "statcheck": Automatically detect statistical reporting inconsistencies to increase

reproducibility of meta-analyses. *Research Synthesis Methods*, *11*(5), 574–579. https://doi.org/10.1002/jrsm.1408

Ren, Z. H. (2012). *ACT-based computerized treatment for depression: Effects, optimal matching, and mechanisms*. Central China Normal University.

Situ, Y. Y. (2022). Effect of acceptance and commitment therapy on adolescent patients with depression. *Chin Foreign Med Res*, *20*(3), 97–100. https://doi.org/10.14033/j.cnki.cfmr.2022.03.026

The Cochrane Collaboration. (n.d.). *Review Manager (RevMan)* [Computer software]. revman.cochrane.org

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically Supported Treatment: Recommendations for a New Model. *Clinical Psychology: Science and Practice*, *22*(4), 317–338. https://doi.org/10.1111/cpsp.12122

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Wang, Y.-P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, *35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Wilkinson, J., Heal, C., Antoniou, G. A., Alfirevic, Z., Avenell, A., Barbour, V., Brown, N. J. L., Carlisle, J., Dicker, P., Dumville, J., Grey, A., Gurrin, L. C., Hayden, J. A., Heathers, J., Hunter, K. E., Lasserson, T., Lam, E., Lensen, S., Li, T., ... Kirkham, J. J. (2023). *Protocol for the development of a tool (INSPECT-SR) to identify problematic randomised controlled trials in systematic reviews of health interventions* (p. 2023.09.21.23295626). medRxiv. https://doi.org/10.1101/2023.09.21.23295626

Zemestani, M., & Mozaffari, S. (2020). Acceptance and commitment therapy for the treatment of depression in persons with physical disability: A randomized controlled trial. *Clinical Rehabilitation*, *34*(7), 938–947. https://doi.org/10.1177/0269215520923135

Zhao, B., Wang, Q., Wang, L., Chen, J., Yin, T., Zhang, J., Cheng, X., & Hou, R. (2023). Effect of acceptance and commitment therapy for depressive disorders: A meta-analysis. *Annals of General Psychiatry*, *22*(1), 34. https://doi.org/10.1186/s12991-023-00462-1