

Adjusting for Publication Bias Reveals Evidence Against Social Comparison As a Behaviour Change Technique Across the Behavioural Sciences

František Bartoš^{1,2*}, Zuzana Iršová², Tomáš Havránek^{2,3}, Eric-Jan Wagenmakers¹

¹ Department of Psychological Methods, University of Amsterdam

² Institute of Economic Studies, Faculty of Social Sciences, Charles University

³ Meta-Research Innovation Center at Stanford

Correspondence concerning this article should be addressed to:

František Bartoš,

University of Amsterdam, Department of Psychological Methods,
Nieuwe Achtergracht 129-B, 1018 VZ, Amsterdam, The Netherlands
f.bartos96@gmail.com

Word count: 858

Social comparison is increasingly used as a behaviour change technique (SC-BCT) yet its true efficacy remains unclear. To address this concern, Hoppen and colleagues¹ performed a comprehensive meta-analysis of 79 randomized controlled trials involving over 1.3 million participants. Hoppen and colleagues¹ “found evidence supporting the efficacy of SC-BCTs in shaping behaviour in the desired direction”. However, their own risk of bias assessment revealed that “most trials ($k = 66$; 84%) had some concern of bias regarding selection of the reported result(s) given that preregistrations and prespecified analysis protocols were rare” and they noted that “it remains unknown to what extent the present results are affected by publication bias”. In our re-analysis, we demonstrate that publication bias exaggerated the evidence in favor of SC-BCT to such a degree that once this bias is properly adjusted for, the effect disappears entirely. In fact, the data show moderate evidence *against* the presence of an effect.

For any meta-analysis it is crucial to report estimates that have been properly adjusted for publication bias—the suppression of non-significant and non-conforming results—, since this bias can inflate effect sizes and produce spurious evidence in favor of an effect^{2,3}. Hoppen and colleagues¹ did test for publication bias using Egger’s regression, and they applied trim and fill publication bias adjustments in cases where Egger’s test was statistically significant. For comparisons with passive control conditions, Egger’s test failed to reject the null hypothesis of no publication bias; for comparisons with active controls, Egger’s test did indicate publication bias, but the trim and fill adjustment had little impact on the estimated effect size. Unfortunately, Egger’s test and trim and fill are known to perform poorly in detecting^{4,5} and adjusting^{2,6,7} for publication bias in the presence of heterogeneity, a scenario that applies to the meta-analysis by Hoppen and colleagues¹.

We re-examine the Hoppen and colleagues¹ data using robust Bayesian meta-analysis⁸. RoBMA accounts for publication bias by combining a set of selection models⁹ and PET-PEESE¹⁰, a meta-regression correction technique based on the funnel plot, via Bayesian model-averaging¹¹. These methods have been shown to greatly outperform publication-bias unadjusted meta-analyses in both simulation studies^{4,5,12} and real data scenarios^{2,12}. Furthermore, Bayesian model averaging allows RoBMA to place more weight on models that predict the data well, making it more robust to model

misspecification and directly evaluate the evidence for the presence vs. absence of the effect, heterogeneity, and publication bias via Bayes factors¹³ (rather than absence of evidence with non-significant p-values).

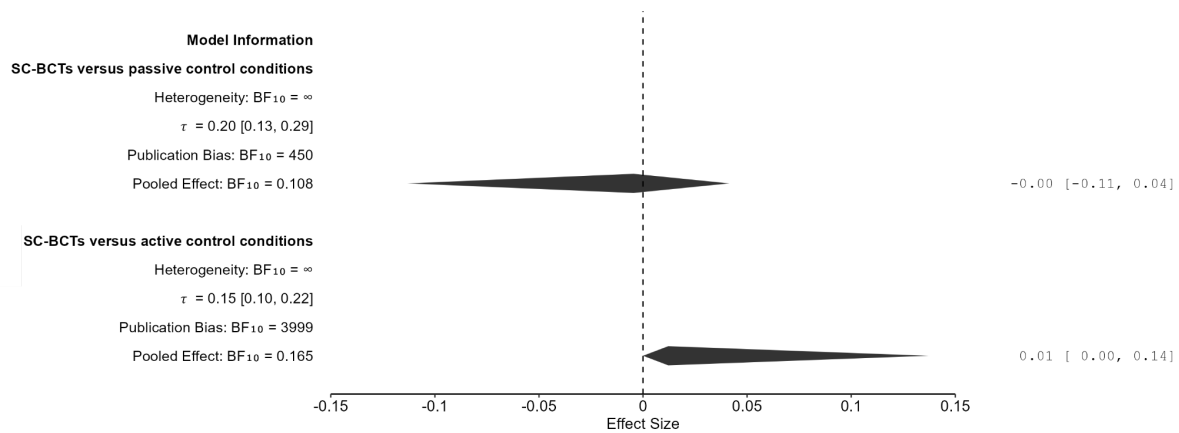
Figure 1 shows the RoBMA meta-analytic publication bias-adjusted estimates for the efficacy of SC-BCT in comparison to passive and active control conditions. After adjusting for publication bias, the efficacy of SC-BCT in comparison to passive control condition decreases from $g = 0.17$, 95% CI [0.11, 0.23] to $g = 0.00$, 95% CI [-0.11, 0.04] and the efficacy of SC-BCT in comparison to active control conditions decreases from $g = 0.23$, 95% CI [0.15, 0.31] to $g = 0.01$, 95% CI [0.00, 0.14]. In essence, the efficacy of SC-BCT completely disappears once publication bias adjustment is performed. Moreover, the evidence for the presence of the effects turns into evidence against the effects, with the data presenting moderate evidence against the effect in comparisons of SC-BCT with passive conditions, $BF_{10} = 0.108$ (i.e., $BF_{01} = 9.3$), and moderate evidence against the effect in comparisons of SC-BCT with active conditions, $BF_{10} = 0.165$ (i.e., $BF_{01} = 6.1$). This means that the observed data are about 9.3 and 6.1 times as likely to occur under the models that assume the absence of the effects than under the models that assume their presence. This reversal in conclusions results from the publication bias detected by and adjusted for by RoBMA; RoBMA shows extreme evidence for the presence of publication bias in both the comparison of SC-BCT with passive control conditions, $BF_{10} = 450$, and active control conditions, $BF_{10} = 3,999$.

Table 1 summarizes the results of a sensitivity analysis with a set of selection models and PET-PEESE individually for all main results reported by Hoppen and colleagues¹. This sensitivity analysis of the short-term efficacy demonstrates that while some of the results differ based on the publication bias adjustment selected, the conclusions are leaning towards the absence of SC-BCT efficacy. The sensitivity analysis of the follow-up effects is much more complicated since the low number of estimates limits the feasibility of publication bias adjustment. Importantly, the individual models included in the sensitivity analysis should not be overinterpreted in isolation, as some might be poorly suited for data with substantial heterogeneity and would receive little weight in the Bayesian model-averaging approach.

Our reanalysis suggests that the apparent efficacy of SC-BCT in shaping behaviour in the desired direction reported by Hoppen and colleagues¹ is largely a byproduct of publication bias. Once publication bias is properly adjusted for, the efficacy of SC-BCT disappears and the data show moderate evidence against the presence of the effect. These findings align with previous publication bias corrected estimates of short term behavioural interventions^{14,15}, which raises the question about the feasibility and effectiveness of simple behavioural change techniques.

Figure 1

Correcting for Publication Bias Shows Evidence Against the Efficacy of Social Comparison as a Behaviour Change Technique



Note. RoBMA model-averaged posterior mean effect size estimates with 95% credible intervals and Bayes factors for the presence of the effect for each outcome. BF_{10} quantifies evidence for the alternative hypothesis. BF_{10} larger than 1 corresponds to evidence in favour of the alternative hypothesis, and BF_{10} lower than 1 corresponds to evidence in favour of the null hypothesis (evidence for the alternative hypothesis can be obtained by inverting the Bayes factor; $BF_{10} = 1/BF_{01}$). As a rule of thumb, Bayes factors between 3 and 10 indicate moderate evidence, and Bayes factors larger than 10 indicate strong evidence. Figure from JASP.

Table 1

Sensitivity Analysis Summarizing the Evidence For and Against the Efficacy of Social Comparison as a Behaviour Change Technique Across Publication Bias Adjustment Methods

	RoBMA	Selection Model (step at $p = 0.025$)	Selection Model (step at $p = 0.025$ and 0.50)	PET-PEESE
Post-intervention results: short-term efficacy				
SC-BCTs versus passive control conditions	0.00 [-0.11, 0.04] $BF_{10} = 0.108$	0.20 [0.10, 0.29] $p < 0.001$	-0.05 [-0.37, 0.28] $p = 0.778$	0.21 [0.14, 0.28] $p < 0.001$
SC-BCTs versus passive control conditions (outlier-adjusted)	-0.01 [-0.17, 0.02] $BF_{10} = 0.122$	0.20 [0.11, 0.29] $p < 0.001$	-0.06 [-0.40, 0.28] $p = 0.734$	0.21 [0.14, 0.29] $p < 0.001$
SC-BCTs versus active control conditions	0.01 [0.00, 0.14] $BF_{10} = 0.165$	0.27 [0.13, 0.41] $p < 0.001$	0.15 [-0.07, 0.36] $p = 0.175$	0.03 [-0.00, 0.06] $p = 0.058$
SC-BCTs versus active control conditions (outlier-adjusted)	0.00 [0.00, 0.15] $BF_{10} = 0.199$	0.25 [0.13, 0.38] $p < 0.001$	0.14 [-0.05, 0.33] $p = 0.150$	0.03 [0.00, 0.06] $p = 0.051$
Follow-up results: long-term efficacy				
SC-BCTs versus passive control conditions	0.03 [0.00, 0.12] $BF_{10} = 0.638$	0.10 [0.06, 0.13] $p < 0.001$	0.10 [0.07, 0.14] $p < 0.001$	0.09 [0.03, 0.15] $p = 0.014$

SC-BCTs versus passive control conditions (outlier-adjusted)	0.06 [0.00, 0.12] BF ₁₀ = 1.928	0.10 [0.07, 0.14] <i>p</i> < 0.001	0.11 [0.07, 0.14] <i>p</i> < 0.001	0.11 [0.07, 0.14] <i>p</i> < 0.001
SC-BCTs versus active control conditions	0.00 [0.00, 0.00] BF ₁₀ = 0.030	0.71 [-0.02, 1.43] <i>p</i> = 0.056	Non-estimable	0.01 [-0.00, 0.03] <i>p</i> = 0.155

Note. RoBMA model-averaged posterior mean effect size estimates with 95% credible intervals and Bayes factors for the presence of the effect for each outcome. Selection models (either one or two-step selection on one-sided *p*-values) and PET-PEESE with effect size estimate with 95% confidence interval and *p*-value against the null hypothesis of no effect.

Data & Materials

The data and analysis script are available at <https://osf.io/rj2g5/>.

Acknowledgements

František Bartoš and Zuzana Iršová acknowledge support from the Czech Science Foundation (grant 23-05227M). Tomáš Havránek acknowledges support from the Czech Science Foundation (grant 24-11583S) and from the Institute for Research on the Socioeconomic Impact of Diseases and Systemic Risks (grant LX22NPO5101), funded by the European Union–Next Generation EU.

Competing interests

Authors declare no competing interests.

Author Contribution

FB analysed the data and wrote the first draft of the manuscript, all authors edited and approved the final version of the manuscript.

References

1. Hoppen, T., Cuno, R., Nelson, J., Lemmel, F., Schlechter, P., & Morina, N. (2025). Meta-analysis of randomized controlled trials examining social comparison as a behaviour change technique across the behavioural sciences. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-025-02209-2>
2. Kvarven, A., Strømmand, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423-434. <https://doi.org/10.1038/s41562-019-0787-z>
3. Bartoš, F., Maier, M., Wagenmakers, E. J., Nippold, F., Doucouliagos, H., Ioannidis, J. P., ... & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods*, 15(3), 500-511. <https://doi.org/10.1002/jrsm.1703>
4. Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research. *Zeitschrift für Psychologie*. 4(227), 261-279. <https://doi.org/10.1027/2151-2604/a000386>
5. Stanley, T.D., Doucouliagos H., Ioannidis, J.P.A., & Carter, E. (2021). Detecting publication selection bias through excess statistical significance. *Research Synthesis Methods*, 12: 776-795. <https://doi.org/10.1002/jrsm.1512>.
6. 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>

7. Hong, S., & Reed, W. R. (2021). Using Monte Carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, 12(2), 192-215. <https://doi.org/10.1002/jrsm.1467>
8. Maier, M., Bartoš, F., & Wagenmakers, E. J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, 28(1), 107-122. <https://doi.org/10.1037/met0000405>.
9. Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419-435. <https://doi.org/10.1007/BF02294384>
10. Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60-78. <https://doi.org/10.1002/jrsm.1095>.
11. Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, 14(4), 382-417. <https://doi.org/10.1214/ss/1009212519>
12. Bartoš, F., Maier, M., Wagenmakers, E. J., Doucouliagos, H., & Stanley, T. D. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, 14(1), 99-116. <https://doi.org/10.1002/jrsm.1594>
13. Jeffreys, H. (1939). *Theory of probability (1st ed.)*. Oxford, UK, Oxford University Press.
14. Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, 119(31), e2200300119. <https://doi.org/10.1073/pnas.2200300119>
15. Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczel, B., ... & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences*, 119(31), e2200732119. <https://doi.org/10.1073/pnas.2200732119>