



No evidence for nudging after adjusting for publication bias

Maximilian Maier^{a,1,2}, František Bartoš^{b,1}, T. D. Stanley^{c,d}, David R. Shanks^a, Adam J. L. Harris^a, and Eric-Jan Wagenmakers^b

Thaler and Sunstein's "nudge" (1) has spawned a revolution in behavioral science research. Despite its popularity, the "nudge approach" has been criticized for having a "limited evidence base" (e.g., ref. 2). Mertens et al. (3) seek to address that limitation with a timely and comprehensive metaanalysis. Mertens et al.'s headline finding is that "choice architecture [nudging] is an effective and widely applicable behavior change tool" (p. 8). We propose their finding of "moderate publication bias" (p. 1) is the real headline; when this publication bias is appropriately corrected for, no evidence for the effectiveness of nudges remains (Fig. 1).

Mertens et al. (3) find significant publication bias, through Egger regression. Their sensitivity analysis (4) indicates that the true effect size could be as low as $d = 0.08$ (if publication bias is severe). Mertens et al. argue that severe publication bias is only partially supported by the funnel plot and proceed largely without taking publication bias into account in their subsequent analyses. However, the reported Egger coefficient ($b = 2.10$) is "severe" (5).

A newly proposed bias correction technique, robust Bayesian metaanalysis (RoBMA) (6), avoids an all-or-none debate over whether or not publication bias is "severe." RoBMA simultaneously applies 1) selection models that estimate relative publication probabilities (7) and 2) models of the relationship between effect sizes and SEs [i.e., Precision Effect Test and Precision Effect Estimate with Standard Error (6, 8, 9)]. Multimodel inference is then guided mostly by those models that predict the observed data best (6, 9, 10). RoBMA makes multimodel inferences about the presence or absence of an effect, heterogeneity, and publication bias (6, 9).

Table 1 compares the unadjusted results to the publication bias-adjusted results.* Since publication bias-corrected three-level selection models are computationally intractable, we analyzed the data in two ways: 1) ignoring the three-level structure (column 2) and 2) using only the most precise estimate from studies with multiple results (column 3). Strikingly, there is an absence of evidence for an overall effect and evidence against an effect in the "information" and "assistance" intervention categories, whereas the evidence is undecided for "structure" interventions. When using only the most precise estimates, we further find evidence against an effect in most of the domains, apart from "other," "food," and "prosocial" (the evidence is indecisive) and weak evidence for the overall effect.† However, all intervention categories and domains apart from "finance" show evidence for heterogeneity, which implies that some nudges might be effective, even when there is evidence against the

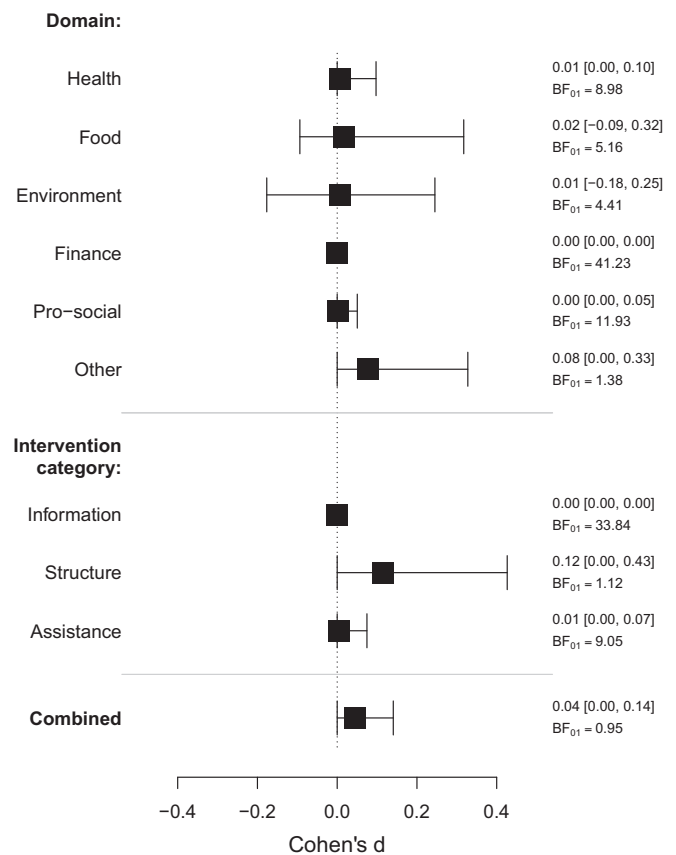


Fig. 1. RoBMA_{PSMA} model-averaged posterior mean effect size estimates with 95% credible intervals and Bayes factors for the absence of the effect for the combined sample or split by either the domain or intervention category (ignoring the clustering of SEs). BF₀₁ quantifies evidence for the null hypothesis. BF₀₁ larger than one corresponds to evidence in favor of the null hypothesis, and BF₀₁ lower than one corresponds to evidence in favor of the alternative hypothesis (evidence for the alternative hypothesis can be obtained by reciprocating the Bayes factor; BF₁₀ = 1/BF₀₁). As a rule of thumb, Bayes factors between 3 and 10 indicate moderate evidence, and Bayes factors larger than 10 indicate strong evidence.

Author affiliations: ^aDepartment of Experimental Psychology, University College London, London WC1H 0AP, United Kingdom; ^bDepartment of Psychological Methods, University of Amsterdam, Amsterdam, 1018 WS, The Netherlands; ^cDeakin Laboratory for the Meta-Analysis of Research, Deakin University, Burwood, VIC 3125, Australia; and ^dDepartment of Economics, Deakin University, Burwood, VIC 3125, Australia

Author contributions: M.M. and F.B. designed research; M.M. and F.B. performed research; M.M. and F.B. analyzed data; and M.M., F.B., T.D.S., D.R.S., A.J.L.H., and E.-J.W. wrote the paper.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹M.M. and F.B. contributed equally to this work.

²To whom correspondence may be addressed. Email: m.maier@ucl.ac.uk.

Published July 19, 2022.

*Our analysis is based on the corrected dataset in ref. 12.

†We also reanalyzed the data by including only models of selection for statistical significance, confirming our results.

Table 1. Comparison of unadjusted and adjusted effect size estimates for all studies and for subsets of studies based on different categories or domains

	Random effects	RoBMA _{PSMA}	RoBMA _{PSMA} (precise)
Combined	0.43 [0.38, 0.48] t(333) = 16.51	0.04 [0.00, 0.14] BF ₀₁ = 0.95	0.11 [0.00, 0.24] BF ₀₁ = 0.31
Intervention category			
Information	0.25 [0.19, 0.30] t(88) = 8.79	0.00 [0.00, 0.00] BF ₀₁ = 33.84	0.00 [0.00, 0.07] BF ₀₁ = 10.57
Structure	0.58 [0.50, 0.66] t(186) = 13.93	0.12 [0.00, 0.43] BF ₀₁ = 1.12	0.23 [0.00, 0.49] BF ₀₁ = 0.33
Assistance	0.22 [0.15, 0.29] t(65) = 6.42	0.01 [0.00, 0.07] BF ₀₁ = 9.05	0.01 [0.00, 0.12] BF ₀₁ = 8.00
Domain			
Health	0.31 [0.22, 0.39] t(64) = 7.03	0.01 [0.00, 0.10] BF ₀₁ = 8.98	0.02 [0.00, 0.19] BF ₀₁ = 3.53
Food	0.66 [0.52, 0.81] t(81) = 9.01	0.02 [-0.09, 0.32] BF ₀₁ = 5.16	0.27 [0.00, 0.64] BF ₀₁ = 0.55
Environment	0.48 [0.37, 0.58] t(56) = 9.16	0.01 [-0.18, 0.25] BF ₀₁ = 4.41	0.00 [-0.44, 0.34] BF ₀₁ = 3.05
Finance	0.23 [0.15, 0.31] t(34) = 6.08	0.00 [0.00, 0.00] BF ₀₁ = 41.23	0.00 [0.00, 0.00] BF ₀₁ = 30.95
Prosocial	0.32 [0.22, 0.42] t(38) = 6.36	0.00 [0.00, 0.05] BF ₀₁ = 11.93	0.05 [0.00, 0.27] BF ₀₁ = 1.89
Other	0.40 [0.29, 0.50] t(55) = 7.66	0.08 [0.00, 0.33] BF ₀₁ = 1.38	0.04 [-0.22, 0.40] BF ₀₁ = 2.45

First column: Random effects metaanalysis estimates with 95% CI based on clustered SEs, all *P* values < 0.001. Second and third columns: RoBMA_{PSMA} model-averaged posterior mean effect size estimates with 95% credible intervals and Bayes factor for the presence of the effect ignoring the clustering of SEs or using the most precise estimates (precise). Results differ slightly from the moderator analysis presented in the article because we analyzed each subfield separately to allow 1) testing for the presence of the effect in each category/domain in the Bayesian framework, and 2) publication bias to operate differently in different subdomains.

mean effect. Finally, we find strong evidence for publication bias across all subdomains (BF_{pb} > 10), apart from food, when using only the most precise estimates (BF_{pb} = 2.49).

We conclude that the “nudge” literature analyzed in ref. 3 is characterized by severe publication bias. Contrary to Mertens et al. (3), our Bayesian analysis indicates that,

after correcting for this bias, no evidence remains that nudges are effective as tools for behaviour change.

Data Availability. Data and analysis script are available in ref. 11.

ACKNOWLEDGMENTS. We thank Mertens et al. for sharing well-documented data and code.

1. R. H. Thaler, C. R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press, 2008).
2. Y. Lin, M. Osman, R. Ashcroft, Nudge: Concept, effectiveness, and ethics. *Basic Appl. Soc. Psych.* **39**, 293–306 (2017).
3. S. Mertens, M. Herberz, U. J. J. Hahnel, T. Brosch, The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2107346118 (2022).
4. J. L. Vevea, C. M. Woods, Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychol. Methods* **10**, 428–443 (2005).
5. C. Doucouliagos, T. D. Stanley, Theory competition and selectivity: Are all economic facts greatly exaggerated? *J. Econ. Surv.* **27**, 316–339 (2013).
6. M. Maier, F. Bartoš, E. J. Wagenmakers, Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychol. Methods*, 10.1037/met0000405 (2022).
7. J. L. Vevea, L. V. Hedges, A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* **60**, 419–435 (1995).
8. T. D. Stanley, H. Doucouliagos, Meta-regression approximations to reduce publication selection bias. *Res. Synth. Methods* **5**, 60–78 (2014).
9. F. Bartoš, M. Maier, E.-J. Wagenmakers, H. Doucouliagos, T. D. Stanley, No need to choose: Model-averaging across complementary publication bias adjustment methods. *Evidence Synthesis Methods*, in press.
10. J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: A tutorial. *Stat. Sci.* **14**, 382–417 (1999).
11. S. Mertens, M. Herberz, U. J. J. Hahnel, T. Brosch, mhhb_nma_data_corrected.csv. Open Science Framework. <https://osf.io/ubt9a/>. Accessed 3 May 2022.
12. M. Maier et al., Code and data for analyses in “No evidence for nudging after adjusting for publication bias.” Open Science Framework. <https://osf.io/svz6e/>. Deposited 6 January 2022.