Datasets included in BEAR
Appendix

2026-01-19

Contents

Arel-Bundock et al 2022 (political science)

Reference: Arel-Bundock et al. (2022).

Research question: assess statistical power in political science research

Data availability: replication package (zip file with date 20241010) can be found at https://osf.io/fgdet. All data are in public domain. We notified one of the authors, Ryan Briggs, who was not aware of any licensing issues.

Data description and source: Authors searched for (and emailed authors of) meta-analysis articles across 141 journals in political science; this led to "a dataset of 16,649 hypothesis tests, grouped in 351 meta-analyses, reported in 46 peer-reviewed meta-analytic articles"

Notes: There is possibly a considerable overlap with Askarov et al dataset, but these two datasets are worth analysing separately given their different focus.

Data processing: using the same code as original authors (replication package runs on makefiles), but skipping some additional processing that they've done, we write a single `estimates.csv` file (3.3 MB). No manipulation of data was done when creating BEAR dataset.

Askarov et al 2023 (economics)

Reference: Askarov et al. (2023).

Research question: Impact of mandatory data sharing on (excessive) statistical significance in economics papers

Data availability: file `Mandatory data-sharing 30 Aug 2022.dta` may be downloaded from https://github.com/anthonydouc/Datasharing/blob/master/Stata/; reproducibility package at https://dataverse.harvard.edu/file.xhtml?fileId=7884702&version=1.0

Data description and source: The paper includes 64,000 economic estimates surveyed by Ioannidis et al. (2017) plus results of "searching through numerous databases, journal websites, and also an email survey of 109 authors known to have produced meta-analyses of economic topics. We cast the widest net possible to capture empirical economic estimates. Thus, some of the papers included in these meta-analyses were originally published in political science and psychology journals. Nevertheless, these meta-studies deal with economics topics." Later they note: "we collect 166,924 parameter estimates from 345 distinct research areas, of which 20,121 were published in 24 of the leading general interest and field journals" The paper mentions 12,521 observations before and 2,537 after data sharing policies in 24 journals; reproducibility dataset has 32 journals (authors removed some smaller ones) and 22,172 rows.

Data processing: We use the same file `Mandatory data-sharing 30 Aug 2022.dta` (11.6 MB) as in the reproducibility package. No additional manipulation of data was done when creating BEAR dataset.

Barnett and Wren

Reference: Barnett and Wren (2019).

Research question: Bias for statistical significance in health and medical journals

Data availability: The dataset Georgescu.Wren.RData may be downloaded from https://github.com/agbarnett/intervals/tree/master/data and also https://github.com/jdwren/ASEC. Data file is `Georgescu.Wren.RData`. Availability: no clear licence given in the repo but the article is CC BY 4.0 in BMJ Open.

Data description and source: The dataset is a collection of confidence intervals for ratio estimates from Medline papers from 1976-2019. Authors scraped (via regular expressions, followed by an independent check using another data mining algorithm, manual checks for 10,000 intervals) 968,000 CIs from abstracts and 350,000 from full texts.

Notes: These are "ratio estimates" like odds ratios, hazard ratios and risk ratios. Note that binary outcomes tend to have (much) lower information than continuous outcomes.

Data processing: We excluded a small fraction of data. First we restricted to 95% confidence intervals only: where `ci.level` was missing we assumed it was 0.95; among intervals with a known `ci.level`, about 0.3% were not 95% and were dropped. We dropped intervals with non-positive width (i.e. `lower >= upper`). We used log scale (we replaced zero or negative lower bounds with a small positive constant; this affected about 0.1% of the sample) and backed out an approximate standard error and point estimate under the usual normal approximation for a 95% CI, setting point estimate to half-point of the interval.

Bartoš et al. 2025 (exercise)

Reference: Bartoš et al. (2025), which builds on another analysis, Singh et al. (2025).

Research question: What is the effect of physical exercise on cognition, memory, and executive function; with extra focus on selective reporting and heterogeneity.

Data availability: Replication data are publicly available via PsyArXiv https://osf.io/preprints/psyarxiv/qr8e2_v1 under Creative Commons By Attribution 4.0 license.

Data description and source: "2,239 effect-size estimates from 215 meta-analyses of randomized controlled trials" was done by Bartoš et al (2025) based on extension of work by Singh et al (2025)

Data processing: z-values were computed as effect size divided by its reported standard error. No filtering or recoding beyond variable renaming was required.

Additional variables used: None.

Brodeur et al. 2024 (economics)

Reference: Brodeur et al. (2024).

Research question: are preregistration and pre-analysis plans associated with reduced p-hacking and publication bias in economics?

Data collection: manually extracted test statistics from RCTs in leading economics journals published 2018-2021. Extraction was performed table by table. Each row corresponds to a single reported test statistic, linked to a paper-level identifier. Unlike most other sources in BEAR, this means we have dozens of estimates per paper: 314 articles with 16,390 estimates

Data availability: replication data file `merged.dta` may be downloaded from https://dataverse.harvard.edu/file.xhtml?fileId=7884702&version=1.0

Data processing: the authors' reported z-statistics were used directly and the reported coefficients and standard errors were retained. No transformation of test statistics or data was required.

Chavalarias et al. 2016 (MEDLINE/PubMed)

Reference: Chavalarias et al. (2016).

Research question: reporting of p-values in the biomedical literature

Data collection: the authors used large-scale text mining of MEDLINE abstracts and PubMed full texts: 4.5mln p-values in 1.6mln MEDLINE abstracts; 3.4mln p-values in 385k PubMed full-text articles.

Data availability: the extracted p-value dataset is publicly available https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6FMTT3 (about 6GB of data) As per the hosting platform: this dataset is released under the Creative Commons Attribution-NonCommercial 4.0 International License. We extracted the relevant columns into `chavalarias.rds` dataset (25 MB).

Data processing: z-values were derived from p-values assuming two-sided tests. P-value truncation operators were processed by collapsing variants such as `<<`, `<<<`, `<=`, `less than`, and `=<` into `<`. We dropped 0.7% of rows where p-values did not have a "plain" format and 0.08% of rows where truncation could not be unambiguously classified.

Additional variables used: source indicator (abstract versus full text).

clinicaltrials.gov snapshot

Trials which reported results at clinicaltrials.gov (National Library of Medicine 2025)

Snapshot date: 12 August 2025

Data source: This is a set of flat file of studies obtained via https://aact.ctti-clinicaltrials.org/ (over 2GBs of data for relevant tables) with relevant columns extracted and joined by us to produce a single table `clinicaltrials.gov_aug2025.rds` (12 MB of data).

Data availability: terms and conditions for ClinicalTrials.gov are at https://clinicaltrials.gov/about-site/terms-conditions. We make distributions of derived effect sizes available in accordance with ToC's. We only use publicly available data and do not imply any endorsement of this work by National Library of Medicine, ClinicalTrials.gov or AACT.

Data processing: extensive. We kept only studies where status was "Completed" and only rows of data where outcome type was "Primary". We set year to the year of `completion_date`.

Both p-values and confidence intervals were reported. See *Standard procedure for dealing with p-values and confidence intervals* below for how we created and chose the z-values. Briefly, we calculated z-values using two separate procedures (based on p-values and confidence intervals) and, if both were available, we made a judgement on which one was more appropriate using a pre-defined rule. We dropped rows only when `z` was missing. Notably, infinite z-values can arise when `p_value` is recorded as exactly 0, and these are retained because they are not missing.

The result of our data processing is `data_cut_with_z.rds`, 2.1MB file.

Cochrane Database of Systematic Reviews (CDSR)

Trials with data available at CDSR, https://www.cochranelibrary.com/ (@ Cochrane Collaboration 2025)

Snapshot date: 20 November 2025

Data collection: using R package `cochrane` (Schwab 2024) we merged data from 8,726 records of reviews in CDSR and merged them into a single file `cdsr_interventions_19nov2025.csv` (63 MB).

Data processing: minimal processing with extensive filtering. Our processed dataset has 39,768 rows, compared to over 800,000 rows of data in the unprocessed dataset.

On processing side, we cleaned up study years, categorised measures (risk ratio, odds ratio, mean difference etc.) and experimental designs, especially an "RCT" flag (based on scanning of abstracts of each review for inclusion criteria). Most studies in CDSR are RCTs but some reviews also allowed quasi-experimental studies, which prevented us from categorising many studies.

Unlike in most of the other datasets in BEAR, we filter CDSR data heavily. First, we only use the outcome and comparison of each review that are coded as "1", as that is most likely the primary outcome and most relevant comparison. This removes over 90% of all rows of data. Secondly, we only use studies with continuous and dichotomous outcomes, removing about 10% of data where effect estimates are based on instrumental variables or based on individual patient data. Lastly, we remove about 5% of data where measure of effect is unknown (retaining: OR, RR, Peto OR, mean difference, standardised mean difference, and risk difference). We also remove some rows where there are zero subjects.

Additional grouping variable: we retained a "source data type" variable to distinguish estimates from published, unpublished, "sought", and mixed data sources.

Costello and Fox 2022; Yang et al 2023, 2024 (ecology and evolution)

Reference: these two datasets come from the same replication package and shared data processing pipeline in Yang et al. 2024 (Yang et al. 2024). Original references for the datasets are Yang et al. 2023 and Costello and Fox 2022 (Yang et al. 2023; Costello and Fox 2022).

Research question: decline effects in ecology (Costello and Fox) and publication bias and performance of empirical research (Yang et al)

Data collection: Costello and Fox look at 466 meta-analyses obtained through searching Web of Science. Reproducibility dataset includes 88,000 rows among 466 analyses in 232 papers. Subsequently, Yang et al removed duplicated studies and zeroes, which meant discarding about 20% of rows. Yang et al collect 102 meta-analyses published 2010-2019 in relevant journals, de-duplicated against Costello and Fox dataset.

Data availability: two datasets may be downloaded from https://github.com/Yefeng0920/replication_EcoEvo_git/. Available under CC BY 4.0 licence.

Data processing: no additional data processing other than cleaning up study years. Note that the replication package for Yang et al does not include meta-analysis indicators.

Additional variables used: we retained a measure variable ("lnRR", "SMD", "Zr" or "uncommon")

EU Clinical Trials Register (EU CTR)

Trials which reported results at the EU Clinical Trials Register, https://www.clinicaltrialsregister.eu/ (European Medicines Agency 2025).

Snapshot date: 18 January 2026

Data collection: we used R package `ctrdata` (Herold 2025) to create a snapshot of EU CTR (`collection = "euctr"`); we downloaded trial records with results only. We stored the downloaded records in a local SQLite database and extracted a fixed set of protocol and results fields plus several fields computed by `ctrdata` (trial phase, sample size, and a "first primary endpoint" p-value/size derived by the package). We saved the extracted flat dataset as `data_euctr_ctgov.rds`, but we only use the EUCTR portion of this combined dataset in BEAR.

Data availability: EUCTR is publicly accessible. The EUCTR legal notice notes that publication on EUCTR does not constitute an endorsement of the information reported.

Data processing: extensive. We use only primary endpoints. We defaulted missing CI levels to 95% and where truncation of p-values was not stated we assumed equality (the latter was the case for about 15% of data). Rest of the derivation of z-values followed the *Standard procedure for dealing with p-values and confidence intervals* described below. We set `year` to record *entry* date (since "completion date" field was not accessible), dropped rows only when z was not finite, and harmonised phase labels to follow the same naming convention as ClinicalTrials.gov.

Additional variables used: we retained phase labels, a measure class (derived from the estimand label), primary endpoint sample size

Head et al 2015 (PubMed papers)

Reference: Head et al. (2015).

Research question: extent of p-hacking and its impact on meta-analyses.

Data collection: PubMed papers that are open access, up to 2014. Authors used regular expressions to extract p-values from text of Abstract and Results, but not tables. Downloaded data has about 220,000 studies and about 2mln rows.

Data availability: dataset available at https://datadryad.org/dataset/doi:10.5061/dryad.79d43 under CC0 1.0 Universal licence.

Data file: `head.rds` (derived from the Dryad files `p.values.csv` and `journal.categories.csv`, with an additional DOI-to-PMID matching table created by us).

Data processing: we followed the same minimal clean-up steps (e.g. removing values found in large supplementary sections) as the original paper and also attached PubMed IDs to the dataset using DOI look-up. We saved a reduced analysis file `head.rds` (9 MB).

When constructing the BEAR dataset, we treated p-values recorded as exactly 0 as truncated and set the z-operator to `>`. We recoded all inequalities to sharp inequalities.

Additional variables used: we use the "Abstract vs Results" variable for grouping

Jager and Leek

Reference: Jager and Leek (2014).

Research question: estimating science-wise false discovery rate.

Data collection: authors used a custom program to extract p-values from scraped PubMed abstracts for papers published in 5 main medical journals 2000-2010. 15,653 p-values are available in 5,322 articles.

Data availability: the data file `pvalueData.rda` is available https://github.com/jtleek/swfdr License for the programs in that repository is GNU GPL, although a license for the dataset is not stated, as far as we are aware.

Data processing: there was only minimal processing. We derived z values from p-values assuming they were two-sided.

Notably, large proportion of p-values is truncated, almost always at 0.0001, 0.001, 0.01, or 0.05. As in all datasets, we retained information on truncation. We also treated p-values recorded as exactly 0 as truncated (z-operator `>`). We created a crude flag for RCTs by searching the paper titles for "randomized", "randomised" and "controlled".

Metapsy (psychotherapy)

Studies included in *Metapsy* (https://www.metapsy.org/), a set of living meta-analytic databases of randomised trials of psychotherapeutic interventions.

Snapshot date: Jan 2026

Data availability: Metapsy is maintained by an international collaboration led by Vrije Universiteit Amsterdam. Databases are available at https://www.metapsy.org/ and via R package `metapsyData` (Harrer, Karyotaki, and Cuijpers 2022). Metapsy FAQ states that all data provided as part of the project are open-access and requires attribution.

Data processing: we merged 20 individual datasets in Metapsy with some cleaning (e.g. to extract study dates from titles). Almost all estimates are Hedges' g, but for one of the databases (`total-response`) we converted log-odds ratios to SMD (dividing by $\Pi/\sqrt{3}$ to scale SD of logistic regression).

Open Science Collaboration 2015

Reference: Open Science Collaboration (2015).

Research question: replications of a quasi-random sample of 100 experiments in psychology.

Data collection: authors of the paper make a comprehensive dataset of results available with their paper. In BEAR we did not use the original studies (since all studies in that set have $p < 0.05$), only replications.

Data availability: the data file `rpp_data.csv` is available as part of repository for the paper https://osf.io/ytpuq/overview under CC0 1.0 Universal license.

Data processing: we derived z values from the replication p-values assuming two-sided p-values (we use the replication p-value column, not the original-study p-value). Because the replication

dataset does not carry direction in a way we use here, z values are unsigned. We treated p-values recorded as exactly 0 as truncated and recorded this as a lower bound on z (z-operator `>`); otherwise we treated p-values as exact (z-operator `=`). We retained the replication sample size as `ss`.

psymetadata, Many Labs 2018 and Nuijten et al (psychology datasets)

Reference: Set of psychology datasets compiled by Rodriguez and Williams (2022). We also disaggregate two datasets: Many Labs 2 project (Klein et al. 2018) a large meta-meta-analysis of intelligence research (Nuijten et al. 2020).

Snapshot date: Jan 2026

Research question: *psymetadata* is an R package that curates multiple open datasets from published meta-analyses in psychology for teaching and demonstrations. Klein et al. (2018) study how replication results vary across labs and settings for a set of classic and contemporary effects in psychology. Nuijten et al. (2020) is a large meta-meta-analysis of intelligence research.

Data availability: 22 datasets are distributed with the `psymetadata` R package (CRAN); after disaggregating Many Labs 2 and intelligence meta-meta-analysis, remaining 20 datasets vary from 65 rows of data to over 1,000. The package is GPL-3; individual datasets originate from the cited source papers.

Data processing: minimal. We used regular expressions to extract year labels from titles. We dropped some invalid rows (missing, non-finite, or non-positive standard errors). For Many Labs we can group by experiment—we use these as `metaid` column.

Additional grouping variable: name of the original psymetadata dataset.

What Works Clearinghouse (education)

Large database of education effect sizes created by Institute of Education Sciences <Institute of Education Sciences (2025)>.

Snapshot date: 17 May 2025

Data availability: flat files are available at https://ies.ed.gov/ncee/wwc/studyfindings. Per permissions and disclaimers section on that website: "Unless stated otherwise, all information on the U.S. Department of Education's IES website ... is in the public domain and may be reproduced, published, linked to, or otherwise used without permission from IES."

Data collection: WWC datasets have 13,054 findings (data-rows) across 1,908 reviews. The dataset includes effect sizes, but not standard errors, as well as p-values reported by the studies and p-values re-calculated by the WWC reviewers.

Data processing: We retained only RCTs and quasi-experimental studies (98% of data altogether). Where available, we preferred effect sizes and p-values calculated by WWC to the ones reported by the studies. We replaced p-values that were equal to zero (this occurred in 3% of the dataset) with truncated p-value (i.e. assumed $p < 10^{-16}$). We calculated z-values from these p-values under assumption of two-sided tests.

Additional grouping variable: "outcome domain" variable (e.g. "academic achievement", "alphabetics")

Standard procedure for dealing with p-values and confidence intervals

Most datasets come with pre-computed effect sizes and standard errors. Some (especially databases of clinical trials and large sets of scraped studies) report p-values. Reporting in clinical trial database is done by investigators and data entry fields are not standardised, requiring more extensive clean-up and a procedure for deriving z-values.

Cleaning up data

For confidence interval levels, we clean them up by setting values above 100 or below 0 to missing, and treating values below 1 as proportions rather than percentages by multiplying by 100. This means that, for example, 0.95 is treated as a 95% confidence interval. For missing values in EU CTR database, we assume 95%.

In clinical trial datasets we classify the effect measure using text matching to a small set of classes (Mean Difference, Odds Ratio, Risk Ratio, Hazard Ratio, Geometric Ratio, Risk Difference, Difference in Percentages, Ratio/Other Ratio, and Other). This string-based mapping is an arbitrary choice. Unstandardised data in these datasets have many hundreds of values.

Confidence intervals

When working with CIs, ratio measures were analysed on the log scale (if and only if lower bound, upper bound, and point estimate were all strictly positive) and non-ratio measures were always kept on the raw scale.

For z-statistics, we computed the critical value using stated CI percent and the reported number of sides (if missing, we defaulted to a two-sided interval). To avoid undefined critical values (e.g. when at 100% or extremely close to it), we bounded the implied α at a small value. We then computed two one-sided standard error estimates from the upper and lower bounds and used their average when both were finite.

We computed a simple symmetry diagnostic for the confidence interval on the chosen scale, defined as the smaller of the two half-widths divided by the larger. We treated intervals as "symmetric enough" only if this ratio exceeded 0.8. This threshold is arbitrary and affects when we trust the CI-derived z-statistic.

We also flagged potentially non-Wald intervals using text matching on type of parameter column, if available, searching for keywords such as median, Hodges, posterior/Bayes, exact, Fieller, bootstrap, and permutation. Any row flagged this way was treated as unreliable for CI-based inference, regardless of its numeric properties.

p-values

The default approach, used for several datasets which report p-values, especially those scraping data from PubMed/Medline, is to treat p-values as two-sided (`z <- qnorm(1 - p/2)`).

For clinical trials databases (clinicaltrials.gov and EU CTR) we sometimes found ambiguity in how p-values were recorded. Sometimes significant results report p-values close to 1, suggesting

that authors reported cumulative distribution, $\Phi(z)$. Therefore for each row we also calculated $2 \cdot \min\{p, 1 - p\}$. When a CI-derived z-statistic was available, we chose whichever p-value led to a z-value that was closest to the CI-derived z.

To avoid dropping rows, we retained unsigned z-statistics when the sign could not be determined. If the sign of effect was known, we applied it to the "p-derived" z; otherwise we used absolute value.

Choice of a single z-value when multiple candidates available

In situations where z can be derived from both CIs and p-values. We choose a single "best" z-statistic per row as follows. We prefer the CI-derived z when it looked like a plausible Wald interval on the chosen scale, meaning: (1) finite z and standard error, (2) positive standard error, (3) not flagged as non-Wald by keyword search in measure label, and (4) either missing symmetry information or symmetry above 0.8. Otherwise we used the p-derived z.

References for datasets used in BEAR

Arel-Bundock, Vincent, Ryan C. Briggs, Hristos Doucouliagos, Marco Mendoza Aviña, and Tom D. Stanley. 2022. "Quantitative Political Science Research Is Greatly Underpowered." *I4R Discussion Paper Series*, no. 6.

Askarov, Zohid, Anthony Doucouliagos, Hristos Doucouliagos, and T D Stanley. 2023. "The Significance of Data-Sharing Policy." *Journal of the European Economic Association* 21 (3): 1191–1226.

Barnett, Adrian Gerard, and Jonathan D Wren. 2019. "Examination of CIs in Health and Medical Journals from 1976 to 2019: An Observational Study." *BMJ Open* 9 (11): e032506.

Bartoš, František, Michaela Lušková, Katerina Bortnikova, Karolína Hozová, Kristína Kantova, Zuzana Irsova, and Tomas Havranek. 2025. "Effect of Exercise on Cognition, Memory, and Executive Function: A Study-Level Meta-Meta-Analysis Across Populations and Exercise Categories." PsyArXiv preprint. https://doi.org/10.31234/osf.io/qr8e2_v1.

Brodeur, Abel, Nikolai M Cook, Jonathan S Hartley, and Anthony Heyes. 2024. "Do Preregistration and Preanalysis Plans Reduce p-Hacking and Publication Bias? Evidence from 15,992 Test Statistics and Suggestions for Improvement." *Journal of Political Economy Microeconomics* 2 (3): 527–61.

Chavalarias, David, Joshua David Wallach, Alvin Ho Ting Li, and John P A Ioannidis. 2016. "Evolution of Reporting p Values in the Biomedical Literature, 1990-2015." *Journal of the American Medical Association* 315 (11): 1141–48.

Cochrane Collaboration. 2025. "Cochrane Database of Systematic Reviews." https://www.cochranelibrary.com/cdsr.

Costello, Laura, and Jeremy W Fox. 2022. "Decline Effects Are Rare in Ecology." *Ecology* 103 (6): e3680.

European Medicines Agency. 2025. "EU Clinical Trials Register." https://www.clinicaltrialsregister.eu/.

Harrer, Mathias, Eirini Karyotaki, and Pim Cuijpers. 2022. *metapsyData: Meta-Analytic Psychotherapy Databases*. https://CRAN.R-project.org/package=metapsyData.

Head, Megan L, Luke Holman, Rob Lanfear, Andrew T Kahn, and Michael D Jennions. 2015. "The Extent and Consequences of p-Hacking in Science." *PLoS Biology* 13 (3): e1002106.

Herold, R. 2025. "Aggregating and Analysing Clinical Trials Data from Multiple Public Registers Using r Package Ctrdata." *Research Synthesis Methods*, 1–33. https://doi.org/10.1017/rsm.2025.10061.

Institute of Education Sciences. 2025. "What Works Clearinghouse Study Findings." https://ies.ed.gov/ncee/wwc/studyfindings.

Jager, Leah R, and Jeffrey T Leek. 2014. "An Estimate of the Science-Wise False Discovery Rate and Application to the Top Medical Literature." *Biostatistics* 15 (1): 1–12.

Klein, Richard A., Michelangelo Vianello, Fred Hasselman, Byron G. Adams, Regan B. Adams, Sinan Alper, Mark Aveyard, Jordan R. Axt, Mayowa T. Babalola, et al. 2018. "Many Labs 2: Investigating Variation in Replicability Across Samples and Settings." *Advances in Methods and Practices in Psychological Science* 1 (4): 443–90. https://doi.org/10.1177/2515245918810225.

National Library of Medicine. 2025. "ClinicalTrials.gov." https://clinicaltrials.gov/.

Nuijten, Michèle B., Marcel A. L. M. Van Assen, Hilde E. M. Augusteijn, Elise A. V. Crompvoets, and Jelte M. Wicherts. 2020. "Effect Sizes, Power, and Biases in Intelligence Research: A Meta-Meta-Analysis." *Journal of Intelligence* 8 (4): 36. https://doi.org/10.3390/jintelligence8040036.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716.

Rodriguez, Josue E., and Donald R. Williams. 2022. "Psymetadata: An r Package Containing Open Datasets from Meta-Analyses in Psychology." *Journal of Open Psychology Data.* https://doi.org/10.5334/jopd.61.

Schwab, Simon. 2024. *Cochrane: Import Data from the Cochrane Database of Systematic Reviews (CDSR).* https://github.com/schw4b/cochrane.

Singh, B., H. Bennett, A. Miatke, D. Dumuid, R. Curtis, T. Ferguson, J. Brinsley, et al. 2025. "Effectiveness of Exercise for Improving Cognition, Memory and Executive Function: A Systematic Umbrella Review and Meta-Meta-Analysis." *British Journal of Sports Medicine* 59 (12): 866–76. https://doi.org/10.1136/bjsports-2024-108589.

Yang, Yefeng, Alfredo Sánchez-Tójar, Rose E O'Dea, Daniel W A Noble, Julia Koricheva, Michael D Jennions, Timothy H Parker, Malgorzata Lagisz, and Shinichi Nakagawa. 2023. "Publication Bias Impacts on Effect Size, Statistical Power, and Magnitude (Type M) and Sign (Type S) Errors in Ecology and Evolutionary Biology." *BMC Biology* 21 (1): 71.

Yang, Yefeng, Erik van Zwet, Nikolaos Ignatiadis, and Shinichi Nakagawa. 2024. "A Large-Scale in Silico Replication of Ecological and Evolutionary Studies." *Nature Ecology & Evolution*, 1–5.