DEPARTMENT OF ECONOMICS AND FINANCE

SCHOOL OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

What Can We Learn From 1000 Meta-Analyses Across 10 Different Disciplines?

Weilun Wu
Jianhua Duan
W. Robert Reed
Elizabeth Tipton

WORKING PAPER

No. 7/2025

Department of Economics and Finance
UC Business School
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 7/2025

What Can We Learn From 1000 Meta-Analyses Across 10 Different Disciplines?

Weilun Wu[1]
Jianhua Duan[2]
W. Robert Reed[1†]
Elizabeth Tipton[3]

April 2025

Abstract: This study describes and analyzes 1000 meta-analyses across ten different disciplines spanning medicine, science, and the social sciences. It highlights significant variations in methodology across disciplines and offers targeted recommendations for enhancing research synthesis. Our analysis reveals discipline-specific differences in the number of studies and estimates per study, types of effect sizes used, and the prevalence of unpublished studies included in meta-analyses. It also examines the extent of effect heterogeneity and the employment of meta-regression to explain this heterogeneity across different fields. Our findings underscore the underutilization of robust meta-analytic methods like three-level models and CR2 clustered standard errors, which are crucial for addressing dependencies among multiple estimates within studies. Finally, we discuss the implications of publication bias and the prevalence of various tests and corrections across disciplines. Our recommendations aim to learn from and apply best practices across all disciplines. This work serves as a resource for researchers conducting their first meta-analyses, as a benchmark for researchers designing simulation experiments, and as a reference for applied meta-analysts aiming to improve their methodological practices.

Keywords: Meta-Analysis, Interdisciplinary comparison, Effect size heterogeneity, Tests for publication bias, Clustering, Meta-analytic estimators, Meta-regression, Statistical software

JEL Classifications: C18, C80, B41

Data Availability Statement: The programming code used to produce the tables and figures in this paper are available at https://osf.io/6dgpn/. The data will be added once the paper is published.

Acknowledgements: We thank Tomas Havranek and participants in the 2023 UCMeta Colloquium, the 2023 MAER-Net Colloquium, and the 2024 Society for Research Synthesis Meetings for helpful feedback and comments. We especially thank Robbie van Aert, Marcel van Assen, Frantisek Bartos, Tomas Havranek (again), Shinichi Nakagawa, and Yefeng Yang for comments on an earlier draft of this paper.

[1] Department of Economics and Finance & UCMeta, University of Canterbury, NEW ZEALAND
[2] Statistics New Zealand, NEW ZEALAND
[3] Statistics for Evidence-Based Policy and Practice Center, Northwestern University, USA

† Corresponding author: W. Robert Reed. Email: bob.reed@canterbury.ac.nz

I. Introduction

Meta-analysis has become an increasingly important tool for synthesizing empirical findings across a wide spectrum of disciplines. This study contributes to the understanding of meta-analytic practices by describing and analyzing 1000 meta-analyses across ten different disciplines spanning medicine, science, and the social sciences. Our primary objective is to assess current meta-analytic practices across these disciplines. These practices include important descriptive information regarding the scope and types of meta-analyses across disciplines, and common methods used in the analysis of meta-analytic data. Here we code key aspects—such as the number of studies, the number of estimates per study, and the degree of effect heterogeneity – and compare these across disciplines. For example, we find that in medicine, meta-analyses are often small (3-5 studies), while in the social sciences they are often large (> 50 studies). We expect that this descriptive information will be particularly useful to statistical and methodological researchers developing methods and conducting simulation studies.

A primary focus in this paper is the examination of methodological approaches used in different fields, and the assessment of the adequacy of these approaches. This involves assessing the alignment between the types of data structures and the analysis methods used. In some cases, we find that these are not aligned, with analysts using methods that – based upon prevailing statistical and methodological knowledge – are inappropriate. For example, we find that in some fields, dependent data are commonly analyzed as if they are independent. By identifying these mismatches, we hope that our paper will provide an opportunity for fields to improve their practices.

To achieve these objectives, we analyzed the first 100 meta-analyses published in November 2021 within each of the following ten disciplines: Anatomy & Physiology; Biology; Business & Economics; Education; Engineering; Environmental Sciences; Medicine;

Pharmacy, Therapeutics & Pharmacology; Psychology; and Public Health. While we aimed for a robust and representative sample within the constraints of available resources, our selection of the first 100 meta-analyses within a specific timeframe provides a snapshot of practices during that period. To avoid oversampling from any single journal, we limited the number of meta-analyses from one journal to ten.

Our study describes a range of key characteristics such as how many studies and estimates were included in meta-analyses and the average number of estimates per study. We record whether unpublished studies were incorporated, the type(s) of effect sizes analyzed, the estimator(s) used (e.g., fixed effects, random effects), and how heterogeneity was reported and quantified. We investigate whether publication bias was assessed, the specific methods employed for that purpose, and whether publication bias was found. We also identify the types of software used to conduct the analyses (e.g., R, Stata, CMA). Based on this analysis, we highlight four areas of practice where there appears to be room for improvement.

This study proceeds as follows. Section II outlines the search process that selected the 1000 meta-analyses for our study. Section III presents descriptive statistics for the meta-analyses from the 10 disciplines in our study. Section IV provides evidence and recommendations for improving four key areas of meta-analytic practice. Section V concludes.

II. Meta-Analysis Search Process

The initial research for this study originated from a blog project exploring disciplinary differences in how meta-analyses are conducted.[1] That preliminary effort sampled 20 meta-analyses from 18 different disciplines. Building on this work, we expanded the scope to include 100 meta-analyses from each of 10 disciplines. We selected this sample size to ensure a robust and representative set of studies within each field, and chose 10 disciplines to enable meaningful cross-disciplinary comparisons. While a larger sample—either in terms of

[1] See https://replicationnetwork.com/2021/05/18/duan-reed-how-are-meta-analyses-different-across-disciplines/

disciplines or meta-analyses per discipline—would have been desirable, practical constraints on time and resources limited the scope of data collection and analysis.

To identify relevant studies, we used the proprietary discovery service Summon[2], available through the university library of the first three authors. Summon is widely adopted by academic, research, and public libraries worldwide. It enables unified searches across multiple databases (e.g., Scopus, JSTOR, PubMed) through a single interface and is particularly valuable for our purposes because it categorizes search results by research discipline. The version we used, Summon 2.0, includes 61 discipline categories spanning the natural sciences, social sciences, applied sciences, and more. For each discipline, we searched using the keyword "meta-analysis," filtering results to include only peer-reviewed journal articles.

Our aim was to identify the 10 disciplines with the highest number of meta-analyses. Here, we acknowledge a degree of disciplinary bias: the home discipline of three authors is Economics, which ranked 13th among the 61 disciplines. To include Economics in our analysis, we combined it with Business (ranked 17th), yielding a sufficiently large set of studies to place the combined category within the top 10. The final list of disciplines in our sample was: Anatomy & Physiology; Biology; Business & Economics; Education; Engineering; Environmental Sciences; Medicine; Pharmacy, Therapeutics & Pharmacology; Psychology; and Public Health.

After selecting the 10 disciplines, we returned to Summon and collected the first 100 meta-analyses published in November 2021 for each field, based on the order in which they appeared. If a discipline did not yield 100 meta-analyses for that month, we extended the search to earlier periods. To avoid overrepresentation from any single journal, we capped the number of meta-analyses per journal at ten. Duplicate entries were removed and replaced with studies

---

[2] For more on Summon, see https://exlibrisgroup.com/products/summon-library-discovery/.

published as close as possible to November 2021. The final list of journals for each discipline is provided in an online appendix. Coding was carried out by multiple teams, with each meta-analysis coded by at least two researchers (and often more), and subsequently reviewed and recoded multiple times by the authors over a multi-year period.

III. A Description of 1000 Meta-Analyses Across 10 Disciplines

<u>Number of studies/Number of estimates</u>. Even a cursory inspection of meta-analyses reveals substantial variation in the number of included estimates and studies. Complicating our task of recording this information was the fact that many studies report multiple meta-analyses—typically for different subsamples—rather than a single, combined meta-analysis. In such cases, we selected the largest meta-analysis for inclusion in our study.

TABLE 1 reports the median values of number of studies and number of estimates for each of the ten disciplines. The disciplines are arranged in ascending order, with the disciplines having the smallest median number of studies at the top of the table. Overall, the median number of studies in a meta-analysis for the full sample is 21, and the median number of estimates is 28. However, there are substantial differences across disciplines.

Pharmacy, Therapeutics & Pharmacology is characterized by the smallest number of studies with a median number of studies of 11, and median number of estimates of 13. Meta-analyses in Business & Economics tend to be the largest, with median values of 53.5 and 163.5, respectively. There are also substantial differences within disciplines. For example, the minimum and maximum number of studies for meta-analyses in Pharmacy, Therapeutics & Pharmacology are 2 and 81. In Business & Economics, minimum and maximum studies are 5 and 613.

This heterogeneity in number of studies and number of estimates is illustrated in FIGURES 1 and 2. The first figure reports boxplots (without outliers) for each of the ten disciplines with respect to number of studies. Superimposed on the boxplots are the average number of studies per meta-analysis. The figure is arranged with smallest median number of studies per meta-analysis at the top of the figure to largest median number of studies at the bottom of the figure. The second figure does the exact same thing, except for estimates per meta-analysis.

Within-discipline heterogeneity is generally increasing in median value. Pharmacy, Therapeutics & Pharmacology; Anatomy & Physiology; Public Health; and Medicine have least heterogeneity in both number of studies and number of estimates and also have the lowest or among the lowest median values of studies and estimates. On the other end, Business & Economics has the greatest heterogeneity in numbers of studies and estimates, and largest median values as well.

One insight from this analysis is that it highlights the difficulty of identifying a "typical" meta-analysis size within any given discipline. Fields such as Pharmacy, Therapeutics & Pharmacology; Anatomy & Physiology; Public Health; and Medicine tend to exhibit greater consistency in the number of included studies and estimates. In contrast, Psychology and Business & Economics show much wider variability. For novice meta-analysts or manuscript reviewers working in these latter fields, it can be difficult to judge what constitutes a normative study size.

A further benefit of the information in TABLE 1 and FIGURES 1 and 2 is its value for informing the design of simulation studies. Researchers conducting simulations to evaluate meta-analytic methods often need to make assumptions about the size and structure of typical datasets. By documenting how these characteristics vary across disciplines, our findings allow researchers to construct simulated datasets that better reflect real-world meta-analytic conditions. This, in turn, enhances the realism, relevance, and generalizability of simulation-based evaluations.

Number of estimates per study. A related characteristic is the number of estimates per study. While the total number of studies and estimates captures overall meta-analysis size, estimates per study draws attention to dependence among observations. For example, the two most common estimators in meta-analysis—fixed effects and random effects—assume that estimated effects are independent, typically reflected in the assumption that each study only has one estimate. If the assumption of independence is violated, it undermines estimator efficiency and compromises the validity of statistical inference. In this section, we investigate whether one estimate per study accurately represents the data structure of most meta-analyses.

TABLE 2 reports the median number of estimates per study for each discipline, sorted in ascending order. In disciplines such as Medicine; Pharmacy, Therapeutics & Pharmacology; Anatomy & Physiology; Public Health; Biology; and Engineering, the median rounds to 1.0 (to one decimal place).[3] At the opposite end, disciplines like Environmental Sciences and Business & Economics have median values above 2.

These values might suggest that most meta-analyses conform to the one-estimate-per-study assumption. However, this conclusion would be reductive and not fully accurate.. The second column of TABLE 2 shows the percentage of meta-analyses in each discipline that include more than one estimate per study. This reveals a more nuanced picture.

For instance, although Engineering has a median of 1.0, approximately half of its meta-analyses include more than one estimate per study. In Environmental Sciences, over 75% do so; in Business & Economics, the figure exceeds 80%. Across the entire sample of 1,000 meta-analyses, 57% contain more than one estimate per study.

FIGURE 3 presents boxplots showing within-discipline heterogeneity in estimates per study, again sorted by median values. Mean values are superimposed for reference. As seen in

[3] Note that the median does not have to a whole number or integer. Say that 100 meta-analyses are rank ordered from smallest to largest number of estimates per study. If the median meta-analysis had 11 estimates from 10 studies, the "median number of estimates per study" would be 1.1.

earlier figures, heterogeneity increases with the median, further highlighting the variability in practice. Clearly, the one-estimate-per-study assumption may be approximately valid in some disciplines—such as Pharmacy, Therapeutics & Pharmacology; Public Health; Medicine; and Anatomy & Physiology—but is inappropriate in others. We discuss the implications of this for estimator efficiency and inference below.

Inclusion of unpublished studies. TABLE 3 highlights another dimension of disciplinary differences: the inclusion of unpublished studies. In disciplines like Environmental Sciences; Medicine; Anatomy & Physiology; Pharmacy, Therapeutics & Pharmacology; and Biology; it is relatively rare to include unpublished studies. Less than 1 in 5 meta-analyses in these disciplines include unpublished studies. In contrast, approximately two-thirds of all meta-analyses in Business & Economics and Psychology include unpublished studies. Overall, about a third (31%) of the meta-analyses in our full sample included unpublished studies.

On the one hand, unpublished studies have not been peer-reviewed, or possibly have been peer-reviewed and rejected. Thus there may be concerns about quality. On the other hand, unpublished studies may be less impacted by pressures to produce novel and statistically significant estimates and, thus, may suffer less from publication bias. This may be especially true when the unpublished literature consists of master and doctoral theses and the authors are not intending to submit their work to academic journals. In Business & Economics, for example, it is common to include unpublished literature and check for differences between published and unpublished studies in subsequent meta-regression or subsample analyses. Without a clear direction for which approach is better, and recognizing differences in the accessibility of unpublished studies, disciplines have evolved to establish their own norms.

Effect sizes. Another dimension that varies widely across disciplines is effect sizes. These are the variables that appear as dependent variables in meta-analyses. For the purposes of classification, we categorized effect sizes into the following groups: Ratios, which include measures such as odds ratios, hazard ratios, and risk ratios; Mean Differences ("Mean Diff"), of which the most common measures are Cohen's d and Hedges g; Prevalence, where the effect size is a percentage or frequency; Correlations ("Corr"), representing correlations and partial correlation coefficients; Fisher's z; and Other, the most common of which are regression coefficients.[4]

FIGURE 4 presents a bar chart that reports overall usage rates of the different types of effect sizes across all disciplines. Most common are Ratio measures, with a little over a third of all meta-analyses using some form of a ratio variable to measure effect size. Closely following are Mean Difference measures. Beyond these two effect sizes, the other categories are roughly used in equal proportions. Note that the sum of the usage rates exceeds 100 percent because some meta-analyses use more than one effect size.

TABLE 4 provides further detail about effect size usage rates by discipline. Disciplines such as Biology; Medicine; and Pharmacy, Therapeutics & Pharmacology most frequently use Ratio-based effect sizes (e.g., odds ratios, hazard ratios, and risk ratios). This could be due to the nature of the data analyzed in these fields where outcomes are commonly binary. In contrast, fields like Anatomy & Physiology, Education, and Psychology show a stronger preference for Mean Difference measures such as Cohen's d and Hedges' g. In these disciplines, it is usual for the outcomes to be continuous and the treatments/interventions to be binary. In Business & Economics, the most frequently used effect size is a correlation. This stems from the fact that it is common in that discipline to combine estimates from regression models that use different outcome and treatment variables. To make the estimates comparable across studies, meta-analyses transform the regression coefficients to partial correlation coefficients (PCCs).

[4] We separated Fisher's *z* from correlations due to several important differences. Although conceptually related, correlations are bounded between –1 and 1 and are directly interpretable, whereas Fisher's *z* is unbounded and not intuitively interpretable in the same way. Most importantly, their standard errors differ: the standard error of *r* depends on the correlation itself, while that of Fisher's *z* depends only on the degrees of freedom.

Lastly, we note that Fisher's z finds it most frequent use in Business & Economics, Education, and Psychology. As we discuss below, some effect sizes, particularly correlations/PCCs, have the problem that the standard error of the estimated effect size is a function of the effect size. When inverse-variance weights are used, this introduces bias in estimating overall mean effects and testing for publication bias. One way to deal with this problem is to transform the original effect sizes to Fisher's z.

Estimators. FIGURE 5 reports the rates at which different meta-analytic estimators are used. We categorized estimators into 7 types: random effects, fixed effects, multivariate/multilevel models ("MVM"), OLS, structural equation models ("SEM"), Bayesian estimators, and Other. While the category MVM includes both multivariate and multilevel models, virtually all the multivariate/multilevel models in our sample were multilevel (Three-Level) models, which is why we combined them.

Perhaps not surprisingly, random effects is far and away the most used meta-analytic estimator. Over 8 out of 10 meta-analyses in our sample used random effects. Fixed effects was a distant second in terms of frequency of use with approximately 1 in 4 meta-analyses using this estimators. All other estimators were used relatively infrequently. As before, we note that the sum of the usage rates within disciplines exceeds 100 percent because many meta-analyses employed more than one estimator.

Detailed usage rates by discipline are reported in TABLE 5. Noteworthy here is that in most disciplines, the only two estimators that matter are random effects and fixed effects. For example, among meta-analyses in Pharmacy, Therapeutics & Pharmacology, 96% used random effects, and 37% used fixed effects. Outside of these, virtually no other estimators were used. Though not as extreme, similar practice was followed in Anatomy & Physiology, Medicine, and Public Health. In this respect, Business & Economics stands out for its diverse usage of estimators. While random effects is the most used estimator in that discipline, substantial percentages of meta-analyses also use MVM, OLS, SEM, Bayesian methods (often Bayesian Model Averaging), and others.

TABLE 6 does a deeper dive into the use of random effects and fixed effects estimators. Specifically, it tracks how often a meta-analysis uses only random effects or only fixed effects. We see that rarely is fixed effects the only estimator used by meta-analyses. In contrast, random effects often is. This is because fixed effects is frequently used to provide an initial analysis of the data before it is subsequently rejected in favour of the random effects model. In other words, even though it is the second most widely used estimator, this does not seem to reflect confidence that it is often viewed as best suited for analyzing the data.

Heterogeneity. One reason for conducting meta-analyses is to combine studies that examine the effect of a common treatment or intervention, with the goal of obtaining a more precise estimate of that effect. A meaningful, and more accurate, estimate is most likely when the studies being pooled are relatively homogeneous. Therefore, it is important to assess the degree to which this is the case.

One measure of effect heterogeneity across studies/estimates within a meta-analysis is tau-squared (τ^2). $\tau^2 = 0$ implies a single population effect, with all observed variability in estimated effects due to sampling error. Positive values of τ^2 indicate heterogeneity in true effects. An advantage of τ is that it is in the same units as the estimated effect, allowing one to compare the extent of heterogeneity on the same scale as the size of the effect. That is also its disadvantage, because it doesn't allow comparison across studies whose estimated effects are in different units. This may be the reason that it is not commonly reported. Column (1) of TABLE 7 indicates that only about a fourth (26%) of meta-analyses in our sample report τ or τ^2.

I-squared addresses this disadvantage. It allows comparison of effect heterogeneity across meta-analyses. It takes values between 0 and 100% (Borenstein et al., 2017) and measures the percent of total variation in estimated effects that is due to effect heterogeneity. High values of I-squared are indicative of a large degree of effect heterogeneity. It is important to emphasize that I-squared is a relative measure of effect heterogeneity, not an absolute one. Two meta-analyses may exhibit the same absolute variation in true effects, but if one includes studies with less precise estimates—leading to greater total variation in observed effects—its I-squared value will be lower. Consequently, differences in I-squared can reflect not only variations in heterogeneity but also differences in sample size and study design.

Column (2) of TABLE 7 shows that I-squared is reported far more frequently than τ^2. Almost three-fourths (72%) of meta-analyses in our sample report an I-squared value, yet here again there are differences across disciplines. For example, only 30% of meta-analyses in Business & Economics do so. However, for most disciplines, reporting I-squared is common practice. Column (3) of TABLE 7 displays median I-squared values, though obviously these median values only apply for the samples of studies that report I-squared. The disciplines in TABLE 7 are sorted from lowest median I-squared value (Pharmacy, Therapeutics & Pharmacology) to highest (Environmental Sciences).

As reference points to interpret the I-squared values in TABLE 7, we note that Higgins et al.'s (2003) seminal article identified an I-squared value of 0.75 as "high" heterogeneity; and the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks et al., 2023) classifies I-squared values above 0.75 as "considerable heterogeneity". Nonetheless, Borenstein et al. (2017) caution against taking this threshold as canonical. While there are notable differences across disciplines, the common result is that all the disciplines in our sample are characterized by substantial heterogeneity.

There are multiple consequences of high heterogeneity. High heterogeneity increases prediction intervals, making average effect sizes less informative as a measure of treatment effects (Borenstein et al. 2017). It also reduces the performance of meta-analytic estimators.

For example, tests and corrections for publication bias become less reliable in the presence of high effect heterogeneity (McShane et al., 2016; Stanley, 2017).

FIGURE 6 provides a more detailed look at the distribution of I-squared within disciplines. Not only do the distributions differ in their central moments, but also in the range of their values. Meta-analyses in Pharmacy, Therapeutics & Pharmacology; Anatomy & Physiology; Medicine; and Biology run the full gamut of values with respect to effect heterogeneity. Other disciplines, especially Business & Economics, are much more concentrated towards the upper half of the scale. Because effect heterogeneity is a key determinant of estimator performance, capturing these differences is important for researchers designing discipline-specific Monte Carlo simulation experiments.

How Researchers Search for Publication Bias—and How Often They Find It. One of the most significant threats to the reliability of meta-analysis is publication bias. Publication bias arises when the studies included in a meta-analysis are not representative of the broader population of research on that subject. This can occur when findings with statistically significant or positive results are more likely to be published, while studies with null or negative findings remain unpublished. As a result, the sample of included studies may be systematically skewed, leading the meta-analysis to yield a distorted and often overly optimistic summary of the empirical literature.

Researchers have developed various approaches to assess publication bias. In a later section, we report which methods for assessing publication bias are most commonly used. In this section, we merely record how often the different disciplines investigate publication bias, and how often they conclude their samples have it. The corresponding results are provided in TABLE 8.

Column (1) of TABLE 8 reports the percent of meta-analyses in each discipline that test for publication bias, where "test" includes qualitative tests such as the funnel plot. Overall, approximately three-fourths (73%) of the meta-analyses assess publication bias. Column (2) shows how often they conclude that publication bias is present. The disciplines in the table are ordered from those where publication bias appears to be most prevalent (Business & Economics) to those where it is found least often (Pharmacy, Therapeutics & Pharmacology).

Given the widespread concern with publication bias, it is notable that some disciplines do not routinely evaluate it. While 9 out of 10 meta-analyses in Psychology investigate publication bias, the rates in Medicine, Environmental Sciences, and Engineering are closer to 6 out of 10. One might think the latter result is partly a function of the number of studies/estimates per meta-analysis, but Anatomy & Physiology has relatively few studies/estimates per meta-analysis (see TABLE 1), and yet 81% of the meta-analyses in that discipline assess publication bias.

It is notable that most assessments for publication bias do not conclude that it is present. Of the 725 meta-analyses that evaluated publication bias, only 31% found evidence of it. In no discipline was the null hypothesis of no publication bias rejected more than 50% of the time. In Engineering and in Pharmacy, Therapeutics & Pharmacology, rejection rates were below 20%.

Of course, caution should be exercised in interpreting these numbers. It is well known that many tests for publication bias suffer from low power, especially when sample sizes are small (Egger et al., 1997; Macaskill et al., 2001; Ioannidis & Trikalinos, 2007). Thus, failure to find evidence of publication bias does not necessarily mean that publication bias is absent. Nevertheless, the results are suggestive that publication bias may not be as serious a problem for some disciplines as it is for others.

Methods used to detect publication bias. In this section we report the methods that researchers employ to detect publication bias. We categorize these as follows:

- Funnel plots (Sterne & Egger, 2001)

- Egger’s regression, including variants of FAT-PET PEESE (Egger et al., 1997, Stanley, 2008)
- Trim and Fill (Duval & Tweedie, 2000)
- Begg and Mazumdar's rank correlation test (Begg & Mazumdar, 1994)
- Fail Safe N test (Rosenthal, 1979)
- Selection model tests (Hedges & Vevea, 1996).
- p-uniform and p-curve tests: (Van Assen et al., 2015; Simonsohn et al., 2014)
- Other

FIGURE 7 shows the prevalence of different methods used to detect publication bias across disciplines. By far the most commonly used approach is the funnel plot, employed in over 60% of all meta-analyses in our sample. The second most frequent method is Egger’s regression test and its variants, used in approximately 46% of cases. Beyond these, the use of alternative methods drops off substantially.

Among the various approaches for assessing publication bias, the funnel plot is distinctive in being a purely qualitative method. For this reason, TABLE 9 separates meta-analyses, by discipline, into three categories: those that rely solely on funnel plots, those that rely solely on other quantitative methods, and those that use both (Columns 1–3, respectively). Each cell in these columns reports two numbers. The top number indicates the unconditional probability that a meta-analysis from a given discipline uses the corresponding category of publication bias assessment. For example, 11% of all meta-analyses in Anatomy & Physiology rely solely on funnel plots. 10% rely solely on alternative quantitative methods, while 59% combine one or more quantitative tests with a funnel plot.

The italicized number below the top number represents the conditional probability—that is, conditional on testing for publication bias, it indicates the percentage of meta-analyses that use each of the three approaches. For example, among the 81 meta-analyses in Anatomy & Physiology that assess publication bias, 14% rely solely on funnel plots, 13% use only quantitative methods other than funnel plots, and 74% combine both approaches. The final column reports the average number of methods employed to assess publication bias. In

Anatomy & Physiology, for instance, meta-analyses that evaluate publication bias use an average of 2.4 methods.

The table shows that 68% of meta-analyses that assess publication bias use a combination of funnel plots and quantitative tests. A few disciplines stand out for their exclusive reliance on funnel plots. In Pharmacy, Therapeutics & Pharmacology, over a third (36%) of meta-analyses that investigate publication bias rely solely on a funnel plot. Similarly, approximately a quarter of meta-analyses in Engineering and Medicine also only use funnel plots to determine whether their samples exhibit publication bias.

TABLE 10 explores in greater detail the different methods used by meta-analyses to detect publication bias. It breaks down the aggregate prevalence rates from FIGURE 7 into discipline-specific numbers. The discipline numbers mostly follow the aggregate rates. Funnel plots and Egger regression tests dominate the other types of assessment methods.

A few disciplines are noteworthy for their disproportionate use of particular methods. For example, almost half of all meta-analyses in Psychology use Trim and Fill. Fail Safe N is mostly used in Business & Economics, Education, Environmental Sciences, and Psychology. And Business & Economics, Education, and Psychology are noteworthy for using a wider variety of methods not employed by other disciplines (e.g., testing for differences between published and unpublished studies; using year of publication to test for time-lag bias (Nakagawa et al., 2021)).

Types of statistical software packages. One other dimension on which meta-analyses differ is the statistical software packages they use. We tracked usage of the following packages.

- R, especially the packages “meta”, “metafor”, “robumeta”, and “dmetar”: free, open-source software environment produced by the R Foundation for Statistical Computing
- Stata, especially the “meta” suite of commands: proprietary software produced by StataCorp
- RevMan (“Review Manager”): free software developed by the Cochrane Collaboration
- CMA (“Comprehensive Meta-Analysis”): proprietary software produced by Biostat
- SPSS: proprietary software produced by IBM
- JASP: free, open-source software produced by researchers at the University of Amsterdam

- Other

FIGURE 8 reports the aggregate usage rate of the different statistical software packages across all disciplines. A wide variety of packages are used. The most common is R, used by over a third of all meta-analyses. The next most common is Stata, used by approximately a quarter of the meta-analyses in our sample. The next most common, in order, are RevMan, CMA, SPSS, and JASP.

TABLE 11 breaks down the overall usage rates by discipline. There are clear differences across the disciplines. While R is the most used statistical package for most disciplines, Stata is preferred in Medicine and Public Health, and RevMan is the most employed package in Pharmacy, Therapeutics, & Pharmacology. R, Stata, and RevMan are approximately evenly used in Astronomy & Physiology and Engineering. We note that researchers wishing to have their systematic reviews included in the Cochrane Database of Systematic Reviews are strongly encouraged to use RevMan.

The choice of statistical package is consequential. For example, the use of point-and-click packages such as CMA, RevMan, and JASP can be restrictive because users are limited by the built-in functionality of those packages. Furthermore, proprietary packages such as CMA and Stata may be problematic for environments that encourage sharing of data and code, as researchers interested in reproducing results from these meta-analyses may not have access to those packages.

IV. Improving Core Practices in Meta-Analysis: Evidence and Recommendations

The previous section documented a wide range of characteristics of meta-analyses. From this analysis, we identified four areas of practice where there appears to be room for improvement. These focus areas were motivated by several key findings: the pervasive presence of effect heterogeneity, the use of correlations as effect sizes, the limited adjustment of effect estimates in response to publication bias, and the infrequent accommodation in estimation procedures for statistical dependence among estimates.

Infrequent use of meta-regression to investigate heterogeneity. TABLE 7 documented that all the disciplines in our sample are characterized by "high" median heterogeneity (Higgins et al., 2003) and "considerable" median heterogeneity (Deeks et al., 2023). In these circumstances, there are potential benefits from investigating the source(s) of this heterogeneity. Two common approaches for this are subsample analysis and meta-regression (Borenstein et al., 2021). We focus on the latter because—unlike subsample analysis—meta-regression enables us to identify the contribution of each characteristic while simultaneously controlling for (i.e., holding constant) the influence of other characteristics.

In meta-analyses, a common goal is to estimate the overall mean of the effect of a treatment or intervention. Meta-regression expands this analysis by adding explanatory variables to the equation. It enables an analysis of how factors related to study design, data characteristics, and estimation methods influence the magnitude of estimated effects, thereby identifying which characteristics contribute most to heterogeneity among these estimates.

Column (1) of TABLE 12 reports the percent of meta-analyses that estimate meta-regressions for each discipline in our sample. In this analysis, to be counted as a meta-regression, the estimated effect size needed to be regressed on some sample, study, or estimation characteristics other than the standard error variable. We did not count univariate, Egger-type regressions as "meta-regressions" because we wanted to focus on the use of meta-regression as a tool for explaining heterogeneity in effect sizes, rather than as a tool to test for publication bias.

The disciplines in TABLE 12 are ordered from least frequent use of meta-regression (Medicine) to most frequent use (Education). The prevalence of meta-regression varies widely across disciplines. Outside the disciplines of Business & Economics, Psychology, and

Education, half or less of all meta-analyses make use of meta-regression to investigate the sources of heterogeneity in estimated effects.

Of these, most are univariate regressions where the estimated effect is regressed on a single sample, study, or estimation characteristic. For example, only 21% of meta-analyses in Medicine estimated a meta-regression. Of these, only 29% used more than one variable to explain the heterogeneity in estimated effects; i.e., a total of 6 meta-analyses (= 21 × 0.29). In fact, other than Business & Economics, univariate meta-regressions comprised the overwhelming majority of meta-regressions across disciplines.

One possible reason for this is that different disciplines may face different benefits and costs in conducting meta-regressions. Column (3) of TABLE 12 reproduces the median I-squared values from TABLE 7. One might think that the disciplines with the most heterogeneity would also have the greatest incentive to conduct meta-analyses. However, the use of meta-regression does not appear to be related to discipline levels of heterogeneity. In contrast, it does appear to be related to sample size.

Columns (4) and (5) of TABLE 12 report median number of studies and median number of estimates for the respective disciplines (reproduced from TABLE 1). There is a clear pattern of a positive relationship between the prevalence of meta-regressions and the size of the meta-analysis samples. This is what one would expect to see as, ceteris paribus, a larger dataset allows one to more precisely attribute heterogeneity to individual factors.

An additional advantage of estimating meta-regressions is their ability to predict effect sizes conditional on specific treatment and outcome characteristics. This allows researchers to estimate how effective a treatment would be under "preferred" or "best-practice" conditions. Meta-regression can also be used in conjunction with risk of bias tools such as RoB 2 (Higgins et al., 2024) and ROBINS-I (Sterne et al., 2024), which help assess the quality and credibility of primary studies. When risk of bias is coded as a study-level moderator and incorporated into the meta-regression, researchers can explore whether and how effect sizes vary systematically with study quality. This can support the identification of treatment effects that are less likely to be inflated by methodological shortcomings—providing a more credible basis for estimating what might happen under optimal conditions.

Despite its potential, the calculation of such “preferred” or “best-practice” estimates from meta-regressions remains significantly underutilized. Among the 1,000 meta-analyses in our sample, only eight reported estimates explicitly derived under best-practice assumptions—seven in Business & Economics and one in Psychology.

In summary, despite substantial heterogeneity being a pervasive characteristic of meta-analyses, the use of meta-regression is relatively limited. In only three disciplines – Business & Economics, Psychology, and Education – did more than 50% of meta-analyses use this tool. Furthermore, the great majority of meta-regressions consist of univariate regressions where the estimated effect sizes are regressed on a single sample, study, or estimation characteristic. This leads to the following recommendations.

RECOMMENDATION #1a: Meta-analysts should increase their use of meta-regression to explore sources of effect heterogeneity and develop “best-practice” estimates whenever sample size and data conditions allow.

RECOMMENDATION #1b: Analysts should go beyond the use of single variable meta-regression and include multiple covariates where possible (Tipton, Pustejovsky, and Ahmadi, 2019).

The use of correlations as effect sizes. As previously shown (cf. FIGURE 4), correlations, including Pearson’s product moment correlation (*r*) and partial correlation correlations (*PCC*), are used in approximately 10 percent of the meta-analyses in our sample. Correlations have the characteristic that the standard error of the estimated effect is a function of the estimated effect.

The standard error of *r* is given by (Jacobs & Viechtbauer, 2017):

(1.a) $SE(r) = \frac{1-r^2}{\sqrt{N-2}}$.

The standard error for PCC is similar (van Aert & Goos, 2023):

(1.b) $SE(PCC) = \frac{1-PCC^2}{\sqrt{df}}$,

where *df* equals the degrees of freedom from the respective regression equation.

At least two problems stem from the fact that the standard error is a mathematical function of the estimated effect size. The first issue concerns the calculation of the overall mean effect. Meta-analyses that employ inverse-variance weighting use the standard error to determine how much weight each estimate receives in the computation of the mean effect. As shown in Equations (1.a) and (1.b), the standard errors of *r* and *PCC* are decreasing functions of the estimated effect size. Specifically, larger absolute values of *r* and *PCC* result in smaller standard error values due to the squared terms in the numerators of both equations. Consequently, these larger estimates receive greater weight in the meta-analysis, which can lead to upwardly biased estimates of the overall mean (Stanley & Doucouliagos, 2023; Hong & Reed, 2024).

A second problem arises when tests for publication bias depend on the standard error, as in Egger-type regressions, Begg and Mazumdar's rank correlation tests, and funnel plots. As shown in Equations (1.a) and (1.b), these tests may detect a spurious relationship between the estimated correlation and its standard error even when no publication bias is present. This can lead to incorrect inferences about the existence of publication bias. A common solution to both problems is to transform correlations to Fisher's *z* values (van Aert, 2023).

Columns (1)-(3) of TABLE 13 focus on estimates of overall mean effect size. Column (1) reports how often a meta-analysis uses a correlation for an effect size. Column (2) then reports how often those same meta-analyses also report Fisher's z estimates. For each and every discipline, Fisher's z estimates are reported less than 15% of the time as a robustness check for correlations. Overall, of the 100 meta-analyses that use correlations as effect sizes, only 7 also report Fisher's z.

An alternative to Fisher's z is to use correlations but weight observations using a sample size-based weight that is independent of the observation's correlation value (e.g., Hunter and Schmidt, 2004). We denote this approach as "N-Weights". Column (3) reports that this is a more common strategy than transforming correlations into Fisher's z. In Business & Economics and Psychology, where the use of correlations are most prevalent, approximately half of all meta-analyses use "N-weights" rather than inverse variance weights.

Columns (4) to (6) of TABLE 13 focus on meta-analyses that use correlations as effect sizes and test for publication bias using Egger regressions, Begg and Mazumdar's rank correlations, and/or funnel plots. As noted above, all of these methods rely on standard errors that are mathematically related to the individual correlations. Column (4) reports the number of meta-analyses that both use correlation-based effect sizes and conduct a publication bias test. Column (5) shows the percentage of these that rely on a standard error-based method (Egger's test, Begg's test, or a funnel plot). Column (6) narrows the focus further, reporting only the percentage that use formal quantitative tests—specifically Egger's or Begg's.

For example, 34 meta-analyses in Business & Economics used a correlation for an effect size and also tested for publication bias. Of these, 79% used standard error-based tests for publication bias, where the standard error was functionally related to the size of the correlation. If we exclude funnel plots from the calculation, the corresponding proportion is 62%. While there are differences across disciplines, it is clear that meta-analyses that use correlations rely heavily on tests for publication bias that depend on the relationship between the estimated effect and the standard error, despite the fact that these are functionally related even in the absence of publication bias.

The above leads us to our next recommendation:

<u>RECOMMENDATION #2</u>: Meta-analyses that use correlations should also include alternative approaches that eliminate the functional dependence of standard errors on estimated effects. One option is to transform correlations to Fisher's z (van Aert, 2023; Stanley et al., 2024). Another option is to proxy correlation standard errors with sample size-based functions that are independent of individual variation in correlations (e.g., Hunter and Schmidt, 2004).

<u>Limited correction for publication bias</u>. Previous sections reported on how frequently meta-analyses tested and found publication bias (cf. TABLE 8) and the methods they used to do so (cf. TABLES 9 and 10). In this section, we report whether meta-analyses attempt to adjust their estimates of overall mean effect after testing for publication bias. Columns (1) and (2) of TABLE 14 reproduce the discipline-level results from TABLE 8, showing how often meta-analyses tested for and found evidence of publication bias. Column (3) reports whether those meta-analyses attempted to adjust their estimates after detecting it.

For most disciplines, the majority of meta-analyses did not go on to correct estimates for publication bias after they found it. For example, 81% of the meta-analyses in Anatomy & Physiology tested for publication bias. 37% of those concluded their samples had publication bias. But only 40% of the 37% of meta-analyses that found publication bias made an adjustment to their original estimates. Only Business & Economics (75%) and Psychology (63%) corrected for publication bias substantially more than half the time they found it.

When applying regression-based methods to correct estimates of the overall mean effect—such as the PET-PEESE approach advocated by Stanley and Doucouliagos (2012)—the meta-analyses in our sample relied almost exclusively on univariate regression models, using the standard error as the sole explanatory variable. Sample, study, or estimation characteristics that might be correlated with the standard error were rarely incorporated. Omitting such variables can distort the estimated impact of publication selection and result in biased corrections of the overall mean effect (Reed, 2023).

Another important factor in correcting for publication bias is heterogeneity. As noted earlier, high levels of heterogeneity are pervasive across all disciplines in our sample (TABLE 7). Under such conditions, regression-based corrections for publication bias become less reliable (Stanley, 2017). There is evidence that selection models may outperform regression methods in these settings (McShane et al., 2016; Hong & Reed, 2021). Despite this, selection models remain relatively underused (cf. TABLE 10).

In light of the above, we offer the following recommendations.

RECOMMENDATION #3a: Meta-analyses should correct for publication bias when tests indicate its presence (Applebaum et al. 2018). As a robustness check, they should also correct for publication bias even when they fail to confirm its presence because power is typically low and tests may not signal it even when it is there. Bias-corrected estimates should be prominently highlighted when reporting results.

RECOMMENDATION #3b: Given concerns about omitted variable bias and high heterogeneity, selection models and multivariate regression models should be considered as robustness checks when correcting publication selection.

RECOMMENDATION #3c: While acknowledging the importance of correcting for publication bias, we call for more research that can guide meta-analysts to methods for how best to do this.

Insufficient accommodation for dependency among estimates. As shown in TABLE 2, many meta-analyses include more than one estimate per study. This likely introduces dependence among the estimated effects, violating the assumption of independence for both fixed and random effects models. There are two general types of dependence - (1) measured on the same people ("correlated effects") and (2) nested in the same study. The first has to do with estimation error while the second has to do with random effects. When dependence is not accounted for properly in a meta-analysis – e.g., by using methods that require independence – the Type I error of the hypothesis tests are incorrect. In general, this results in tests that reject null hypotheses (e.g., $p < .05$) more often than they should.

There are various approaches to addressing this dependence, depending upon its type. When effect sizes are measured on different people but are nested in the same study (cf. (2) above), then a Multilevel Model (MLM) is an appropriate approach. The most common of which is three-level meta-analysis (Van den Noortgate et al., 2013). A difficulty with this approach is that it requires large sample sizes.

Importantly, the MLM approach is only valid when there is no dependence induced from estimation, i.e., from effect sizes measured on the same individuals (cf. (1) above). When this type of dependence arises, a different approach is needed. One approach is Multivariate Meta-analysis (MVMA; Hedges & Olkin, 1984). In this approach, the dependence structure is modelled directly in terms of a variance-covariance matrix. While this approach has been available for a long time, a difficulty with implementation is that it requires estimates of the correlation between effect sizes. These correlations, however, are often not provided in primary studies. When this is the case, this model-based approach can be incorrect, again leading to inflated Type I errors.

An alternative approach to dependency is to prioritize the accuracy of standard error estimation over the efficiency of coefficient estimates and use clustered standard errors. Cluster-Robust Standard Errors (CRSE) can be applied in any regression framework. In the context of meta-analysis, Robust Variance Estimation (RVE) offers a specialized form of CRSE that accounts for the weighting schemes and sample size structures typical of meta-analytic data (Hedges, Tipton, & Johnson, 2010; Tipton, 2015). The theoretical foundation for both CRSE and RVE is asymptotic, where it can be demonstrated that estimators converge to the correct standard errors as the number of studies becomes large.

Various adjustments have been developed to improve the performance of cluster robust standard error estimators when there are small-to-moderate numbers of studies. In regression, the adjusted CRSE estimator is called CR1. CR1 is implemented broadly in regression-oriented packages like Stata. In meta-analysis, the associated standard error estimator is called CR2. CR2 is the default in RVE implementations in both R and Stata.

The CR2 estimator is known as the bias-adjusted, cluster-robust variance estimator or small-sample corrected cluster-robust variance estimator (Bell & McCaffrey, 2002; Tipton, 2015; Tipton & Pustejovsky, 2015; Pustejovsky & Tipton, 2018). It has been shown to have better inference properties when the number of clusters (e.g., studies) is small, while the CR1 estimator has been shown to have inflated Type I errors, even with as many as 50 or 70 studies. Since the CR2 adjustment also performs well when the number of clusters is large, it should generally be preferred to the CR1 estimator (Huang & Li, 2022).

Both the CR1 and CR2 adjustments result in the use of a t-distribution for hypothesis testing. For the CR1 approach, however, these degrees of freedom are based entirely on the number of clusters/studies and are the same for all regression coefficients. The CR2 degrees of freedom, however, are estimated using a Satterthwaite approximation (Satterthwaite, 1946). Tipton (2015) shows that, in general, RVE has valid Type I error when the CR2 degrees of freedom are greater than 4. When they are smaller than 4, the approximation can lead to inflated Type I errors. Accordingly, when the degrees of freedom are smaller than 4, it is suggested to use a higher standard of evidence (e.g., $p < .01$ instead of $p < .05$).

Importantly, these degrees of freedom are a complex function of the number of studies, the variation in the number of estimates per study, and the distribution of the covariate being tested. As a result, in meta-regression, different coefficients can have different associated degrees of freedom. Because of the complexity of the calculation, even when the number of studies is large (e.g., 70), it is possible for these degrees of freedom to be very small, and it is not possible to know – without calculating them – when such small degrees of freedom might occur. It is for this reason that Tipton, Pustejovsky, and Ahmadi (2017) propose that the CR2 method should be the default in all implementations of RVE, as found in the R and Stata robumeta packages.

Our study analyzed how meta-analyses addressed dependence when it arose. To do so, we first recorded if there were dependent effect sizes by noting when the number of effect sizes was larger than the number of studies. If there was dependence (number of effect sizes > number of studies), we then coded how it was addressed. We coded several common approaches: (1) dependence (incorrectly) ignored; (2) a single effect size selected for analysis in each study to induce independence; (3) effect sizes averaged to the study level to remove dependence; (4) use of multilevel modelling (MLM); (5) use of multivariate meta-analysis (MVMA); (6) use of CRSE with CR1; (7) use of RVE with CR2; and (8) other.

Note that meta-analysts could use a combination of these methods (e.g., MLM and RVE). Importantly, while we code the methods used in these meta-analyses, we did not assess if these methods fully accounted for dependence. For example, a study with correlated effects might have handled this appropriately through the use of (2) selection, (3) averaging, or (7) RVE, but might have inappropriately handled this through (4) use of an MLM model (which adjusts for a different type of dependence). However, regardless of approach, we consider any of these methods better than (1) incorrectly ignoring the dependence and proceeding as if the data were independent (e.g., using a conventional random effects model).

TABLE 15 presents the results. Overall, 57% (571/1000) of the meta-analyses included dependent effect sizes (i.e., number of effect sizes > number of studies). Dependent effects were most common in Business and Economics (84%), Environmental Science (78%), Psychology (77%) and Education (69%). However, of the 571 meta-analyses that had dependent effects, only 269 (47%) adjusted for this dependence in their analysis. That is, 53% of meta-analyses with dependent effects treated the data as independent. Some fields adjusted for dependence more often than others. For example, Psychology (77%) and Business and

Economics (73%) were more likely to adjust for dependence; while Pharmacy, Therapeutics, and Pharmacology (10%) and Medicine (15%) were least likely.

TABLE 15 also indicates the types of methods for adjusting for dependence that are commonly used. Overall, selecting (21%) and averaging (30%) effects were common across fields. Multilevel models were the most common approach in Biology (76%) and Medicine (80%). Multivariate models were rarely used (except for 10% in Biology). Finally, CRSE approaches were more common in Business and Economics (39% use CR1) and Education (48% use CR2).

In light of these results, we provide the following recommendations:

RECOMMENDATION #4a: Dependent effects need to be appropriately addressed in meta-analyses. The use of multilevel models, multivariate models, and cluster robust standard errors enable inclusion of all of the effect sizes and should be employed more commonly.

RECOMMENDATION #4b: Meta-analysts using cluster robust standard errors should estimate CR2 standard errors and routinely check whether the Satterthwaite degrees of freedom are greater than or equal to 4. When the degrees of freedom are smaller than 4, higher standards of evidence should be implemented (e.g., considering Type I error of 0.01 instead of 0.05 for null hypothesis rejection). CR1 should rarely, if ever, be used because the CR2 estimator generally produces results as good or better.

V. Conclusion

This study examines 1000 meta-analyses across ten diverse disciplines, revealing substantial methodological differences and suggesting ways to refine meta-analytic techniques. We found significant variations in the number of studies included in meta-analyses and the number of estimates per study; the types of effect sizes and estimators used; the inclusion of unpublished studies; practices regarding the testing and treatment of publication bias; the use of meta-regression; and the types of statistical software packages employed. Our recommendations encourage wider use of Fisher's z as an effect size, greater application of corrective procedures for publication bias, increased use of meta-regression, and wider adoption of sophisticated methods such as multilevel models, multivariate models, and robust variance estimation with

CR2 adjustments to handle data dependencies. It is hoped that this study will serve as a resource for researchers conducting their first meta-analyses, as a benchmark for researchers designing simulation experiments, and as a reference for applied meta-analysts aiming to improve their methodological practices.

REFERENCES

Altman D.G. (1991). Practical statistics for medical research. London: Chapman and Hall.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2), 169-182.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5-18.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis* (2nd ed.). John Wiley & Sons.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.

Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology & Community Health*, 62(8), 752-758.

Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. (https://training.cochrane.org/handbook/current/chapter-10

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629-634.

Elff, M., Heisig, J. P., Schaeffer, M., & Shikano, S. (2021). Multilevel analysis with few clusters: Improving likelihood-based methods to provide unbiased estimates and accurate inference. *British Journal of Political Science*, 51(1), 412-426.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, 21(4), 299-332.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.

Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Hong, S., & Reed, W. R. (2021). Using Monte Carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, *12*(2), 192-215.

Hong, S., & Reed, W. R. (2024). Meta-analysis and partial correlation coefficients: A matter of weights. *Research Synthesis Methods*, *15*(2), 303-312.

Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods*, 1-19.

Hunter, J. E., & Schmidt, F. L. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. *Journal of Applied Psychology*, *79*(2), 171.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3), 245-253.

Jacobs, P., & Viechtbauer, W. (2017). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Research Synthesis Methods*, *8*(2), 161-180.

Lin, L. (2020). Hybrid test for publication bias in meta-analysis. *Statistical Methods in Medical Research*, 29(10), 2881-2899.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, 20(4), 641-654.

MacKinnon, J. G. (2019). How cluster-robust inference is changing applied econometrics. *Canadian Journal of Economics/Revue Canadienne d'Economique*, 52(3), 851-881.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11(5), 730-749.

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., ... & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*(1), 4-21.

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4), 672-683.

Reed, W.R. (Dec 19, 2023). Univariate PET-PEESE: Why not multivariate PET-PEESE? *MAER-NET*. https://www.maer-net.org/post/univariate-pet-peese-why-not-multivariate-pet-peese

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, 21(4), 37-59.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110-114.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: Genera*l, 143(2), 534.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, 70(1), 103-127.

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, 8(5), 581-591.

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. New York: Routledge.

Stanley, T. D., & Doucouliagos, H. (2023). Correct standard errors can bias meta-analysis. *Research Synthesis Methods*, *14*(3), 515-519.

Stanley, T. D., Doucouliagos, H., Maier, M., & Bartoš, F. (2024). Correcting bias in the meta-analysis of correlations.*Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000662

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046-1055.

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature, *Journal of Clinical Epidemiology,* 53, 1119-1129

Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604-634.

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, *10*(2), 161-179.

van Aert, R. C. (2023). Meta-analyzing partial correlation coefficients using Fisher's z transformation. *Research Synthesis Methods*, *14*(5), 768-773.

van Aert, R. C., & Goos, C. (2023). A critical reflection on computing the sampling variance of the partial correlation coefficient. *Research Synthesis Methods*, *14*(3), 520-525.

Van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, 20(3), 293.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45, 576-594.

TABLE 1
Number of Studies and Estimates in Meta-Analyses

Discipline	Median No. of Studies	Median No. of Estimates
Pharmacy, Therapeutics & Pharmacology	11	13
Anatomy & Physiology	14	19.5
Public Health	15	17.5
Engineering	15.5	20.5
Medicine	16	17.5
Biology	21	25.5
Environmental Sciences	22.5	63.5
Education	28.5	41.5
Psychology	32	62
Business & Economics	53.5	163.5
Overall	**21**	**28**

NOTE: The data in the table are based on 100 meta-analyses for each discipline. As noted, the numbers in the table report median values. A more detailed summary is provided in FIGURES 1 and 2.

TABLE 2
Number of Estimates per Study in Meta-Analyses

Discipline	Median Estimates/Study (1)	% MAs ≥ One Estimate/Study (2)
Medicine	1.0	33%
Pharmacy, Therapeutics & Pharmacology	1.0	39%
Anatomy & Physiology	1.0	46%
Public Health	1.0	47%
Biology	1.0	48%
Engineering	1.0	50%
Psychology	1.2	77%
Education	1.3	69%
Environmental Sciences	2.0	78%
Business & Economics	3.1	84%
Overall	**1.1**	**57%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in Column (2) also provide the counts (e.g., 33% = 33 meta-analyses in the given discipline). As noted, the numbers in the table report median values. A more detailed summary is provided in FIGURE 3.

TABLE 3
Percent of Meta-Analyses Including Unpublished Primary Studies

Discipline	Percent Using Unpublished Studies
Environmental Sciences	9%
Medicine	14%
Anatomy & Physiology	15%
Pharmacy, Therapeutics & Pharmacology	15%
Biology	18%
Engineering	21%
Public Health	29%
Education	56%
Business & Economics	64%
Psychology	68%
Overall	**31%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 9% = 9 meta-analyses in the given discipline).

TABLE 4
Prevalence of Different Effect Sizes

Discipline	Ratio	Mean Diff	Prevalence	Corr	Fishers z	Other
Anatomy & Physiology	37%	51%	8%	2%	4%	2%
Biology	64%	25%	16%	1%	3%	5%
Business & Economics	8%	11%	0%	46%	21%	19%
Education	15%	56%	2%	14%	17%	4%
Engineering	44%	36%	11%	6%	3%	13%
Environmental Sciences	44%	23%	9%	5%	4%	24%
Medicine	56%	20%	34%	0%	3%	4%
Pharmacy, Therapeutics & Pharmacology	54%	40%	18%	0%	2%	1%
Psychology	8%	43%	5%	23%	23%	5%
Public Health	47%	31%	21%	3%	2%	8%
Overall	**38%**	**34%**	**12%**	**10%**	**8%**	**9%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 37% = 37 meta-analyses in the given discipline). Note that the sum of the percentages (counts) across each row are greater than 100% (100 meta-analyses) because studies can use more than one effect size.

TABLE 5
Prevalence of Different Estimators

Discipline	RE	FE	MVM	OLS	SEM	Bayes	Other
Anatomy & Physiology	91%	32%	5%	2%	0%	2%	1%
Biology	83%	22%	16%	5%	1%	2%	1%
Business & Economics	68%	26%	24%	21%	19%	12%	16%
Education	88%	11%	11%	2%	0%	0%	3%
Engineering	78%	30%	1%	13%	1%	1%	3%
Environmental Sciences	64%	23%	16%	15%	1%	1%	4%
Medicine	87%	34%	4%	4%	0%	1%	1%
Pharmacy, Therapeutics & Pharmacology	96%	37%	1%	0%	0%	0%	3%
Psychology	83%	15%	20%	5%	5%	1%	0%
Public Health	92%	34%	4%	5%	0%	0%	2%
Overall	**83%**	**26%**	**10%**	**7%**	**3%**	**2%**	**3%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 92% = 92 meta-analyses in the given discipline). Note that the sum of the percentages (counts) across each row are greater than 100% (100 meta-analyses) because studies can use more than one estimator. RE = Random Effects; FE = Fixed Effects; MVM = Multivariate/Multilevel Models; OLS = Ordinary Least Squares; SEM = Structural Equation Models; Bayes = Bayesian estimation.

TABLE 6
Closer Comparison of Random Effects and Fixed Effects

Discipline	OnlyRE	OnlyFE
Anatomy & Physiology	60%	5%
Biology	57%	0%
Business & Economics	25%	1%
Education	74%	0%
Engineering	50%	4%
Environmental Sciences	42%	1%
Medicine	59%	6%
Pharmacy, Therapeutics & Pharmacology	60%	3%
Psychology	57%	0%
Public Health	56%	1%
Overall	**54%**	**2%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 64% = 64 meta-analyses in the given discipline). OnlyRE reports how many meta-analyses use only one estimator and that estimator is Random Effects. OnlyFE reports how many meta-analyses use only one estimator and that estimator is Fixed Effects.

TABLE 7
Reporting of Effect Heterogeneity and Median I-Squared

Discipline	% Reporting Tau-squared (1)	% Reporting I-squared (2)	Median I-squared (3)
Pharmacy, Therapeutics & Pharmacology	43%	96%	67%
Anatomy & Physiology	40%	88%	69%
Medicine	23%	89%	75%
Biology	20%	66%	77%
Engineering	18%	72%	78%
Public Health	26%	86%	83%
Psychology	27%	71%	84%
Education	35%	76%	86%
Business & Economics	13%	30%	89%
Environmental Sciences	15%	50%	91%
Overall	**26%**	**72%**	**80%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 43% = 43 meta-analyses in the given discipline). The numbers in the second column report median values across the set of meta-analyses in that discipline that report an I-squared value.

TABLE 8
Percent of Meta-Analyses Testing and Finding Publication Bias

Discipline	Testing for Publication Bias (1)	Finding Publication Bias (2)
Business & Economics	74%	43%
Psychology	90%	39%
Anatomy & Physiology	81%	37%
Public Health	69%	35%
Biology	75%	35%
Medicine	61%	33%
Education	85%	29%
Environmental Sciences	58%	24%
Engineering	58%	17%
Pharmacy, Therapeutics & Pharmacology	74%	16%
Overall	**73%**	**31%**

NOTE: The numbers in Column (1) of the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 74% = 74 meta-analyses in the given discipline). The numbers in Column (2) report conditional probabilities. For example, 43% in the first row of the table means that 43% of the 74 meta-analyses in Business & Economics (i.e., 32 of the 74 meta-analyses) that tested for publication bias concluded that there was publication bias.

TABLE 9
Assessing Publication Bias: Funnel Plots and Other Approaches

Discipline	Only Funnel (1)	Other Than Funnel (2)	Funnel + Other (3)	Mean # of Methods (4)
Anatomy & Physiology	11% *(14%)*	10% *(13%)*	59% *(74%)*	2.4
Biology	12% *(16%)*	9% *(12%)*	55% *(72%)*	2.2
Business & Economics	4% *(5%)*	25% *(34%)*	45% *(61%)*	2.3
Education	5% *(6%)*	14% *(17%)*	64% *(77%)*	2.5
Engineering	14% *(24%)*	5% *(9%)*	39% *(67%)*	2.1
Environmental Sciences	10% *(17%)*	15% *(25%)*	34% *(58%)*	2.2
Medicine	15% *(23%)*	11% *(17%)*	38% *(59%)*	1.9
Pharmacy, Therapeutics & Pharmacology	26% *(36%)*	4% *(5%)*	43% *(59%)*	1.9
Psychology	5% *(6%)*	16% *(18%)*	69% *(77%)*	2.6
Public Health	14% *(20%)*	9% *(13%)*	46% *(67%)*	2.2
Overall	**12%** ***(16%)***	**12%** ***(16%)***	**49%** ***(68%)***	**2.2**

NOTE: The top numbers in the first three columns of the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts. Thus 11% =11 meta-analyses in Anatomy & Physiology use only a funnel plot to assess publication bias; 10% = 10 meta-analyses use something besides a funnel plot; and 59% = 59 meta-analyses use a combination of both. Thus 80 meta-analyses in Anatomy & Physiology assess publication bias in some way. The italicized number below the top number is the conditional probability. Conditional on assessing publication bias, it shows what percent of meta-analyses use each of the three approaches. Thus, among all meta-analyses in Anatomy & Physiology that assess publication bias, 14% solely use funnel plots, 13% only use quantitative tests other than funnel plots, and 74% use a combination of both. The last column reports the discipline mean number of tests employed by meta-analyses that assess publication bias. This number includes funnel plots. For example, meta-analyses in Anatomy & Physiology that test for publication bias use an average of 2.4 tests.

TABLE 10
Prevalence of Methods to Assess Publication Bias

Discipline	Funnel	Eggers	TrimFill	Beggs	FailSafe	PUniCurv	Selection	Other
Anatomy & Physiology	70%	56%	28%	26%	8%	0%	0%	2%
Biology	67%	51%	15%	17%	8%	1%	0%	4%
Business & Economics	49%	42%	15%	7%	22%	3%	5%	20%
Education	69%	52%	29%	17%	20%	2%	0%	15%
Engineering	53%	36%	13%	13%	5%	0%	0%	0%
Environmental Sciences	44%	36%	11%	13%	17%	0%	0%	5%
Medicine	53%	36%	15%	13%	2%	0%	0%	2%
Pharmacy, Therapeutics & Pharmacology	69%	45%	4%	20%	1%	0%	0%	0%
Psychology	74%	58%	45%	8%	21%	4%	2%	13%
Public Health	60%	47%	18%	20%	4%	0%	0%	1%
Overall	**61%**	**46%**	**19%**	**15%**	**11%**	**1%**	**1%**	**6%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 70% = 70 meta-analyses in the given discipline). Note that the sum of the percentages (counts) across each row are greater than 100% (100 meta-analyses) because studies can use more than one method to detect publication bias. Funnel = Funnel plot; Eggers = Egger-type regression; TrimFill = Trim and Fill; Beggs = Begg and Mazumdar's rank correlation test; FailSafe = Fail Safe N; Selection = publication bias uses a selection model; PUniCurv = either p-Uniform or p-Curve test for publication bias.

TABLE 11
Prevalence of Different Statistical Packages

Discipline	R	Stata	RevMan	CMA	SPSS	JASP	Other
Anatomy & Physiology	28%	26%	28%	16%	0%	1%	11%
Biology	55%	28%	15%	6%	3%	0%	8%
Business & Economics	33%	12%	2%	9%	2%	0%	12%
Education	38%	19%	11%	26%	1%	1%	6%
Engineering	26%	28%	26%	10%	5%	0%	14%
Environmental Sciences	48%	27%	4%	4%	4%	0%	13%
Medicine	29%	38%	12%	17%	2%	0%	11%
Pharmacy, Therapeutics & Pharmacology	26%	36%	42%	6%	2%	1%	8%
Psychology	54%	5%	1%	29%	5%	1%	4%
Public Health	32%	43%	14%	5%	2%	0%	7%
Overall	**37%**	**26%**	**16%**	**13%**	**3%**	**0%**	**9%**

NOTE: The data in the table are based on 100 meta-analyses for each discipline, so the percentages in the table also provide the counts (e.g., 28% = 28 meta-analyses in the given discipline). Note that the sum of the percentages (counts) across each row can be greater than 100% (100 meta-analyses) because studies sometimes use more than one statistical package.

TABLE 12
Use of Meta-Regression in Meta-Analyses

Discipline	% Estimating Meta-Regressions (1)	% MRs with > 1 Regressor (2)	Median I-squared (3)	Median No. of Studies (4)	Median No. of Estimates (5)
Medicine	21%	29%	75%	16	17.5
Pharmacy, Therapeutics & Pharmacology	31%	32%	67%	11	13
Public Health	31%	29%	83%	15	17.5
Engineering	35%	31%	78%	15.5	20.5
Environmental Sciences	42%	24%	91%	22.5	63.5
Anatomy & Physiology	47%	21%	69%	14	19.5
Biology	49%	27%	77%	21	25.5
Business & Economics	76%	79%	89%	53.5	163.5
Psychology	77%	14%	84%	32	62
Education	81%	40%	86%	28.5	41.5
Overall	**49%**	**35%**	**80%**	**21**	**28**

NOTE: The data in the first column of the table are based on 100 meta-analyses for each discipline, so the percentages also provide the counts (e.g., 21% = 21 meta-analyses in the given discipline). The values in the Column (2) report the percentage of meta-regressions that have more than one regressor. The median I-squared values in Column (3) are reproduced from, and explained in, TABLE 7. The median numbers of studies and estimates in Columns (4) and (5) are reproduced from, and explained in, TABLE 1.

TABLE 13
Correlations, Fisher's z, and Alternative Weights

Discipline	EFFECT SIZES: Correlation & Fisher's z		WEIGHTS	TESTING FOR PUBLICATION BIAS: Egger's (E), Begg's (B), Funnel Plot (F)		
	Count (1)	**Fisher's z (2)**	**N-Weights (3)**	**Count (4)**	**E/B/F (5)**	**E/B (6)**
Anatomy & Physiology	2	0%	0%	0	----	----
Biology	1	0%	100%	1	0%	0%
Business & Economics	46	9%	50%	34	79%	62%
Education	14	14%	21%	11	73%	64%
Engineering	6	0%	17%	2	100%	0%
Environmental Sciences	5	0%	0%	3	33%	33%
Medicine	0	----	----	0	----	----
Pharmacy, Therapeutics & Pharmacology	0	----	----	0	----	----
Psychology	23	4%	48%	21	62%	38%
Public Health	3	0%	33%	3	100%	67%
Overall	**100**	**7%**	**40%**	**75**	**72%**	**52%**

NOTE: Columns (1) reports how many meta-analyses in the respective disciplines have an effect size that is a correlation. Column (2) reports how often the respective meta-analyses also use a Fisher's z effect size as a robustness check. Column (3) reports how often meta-analyses that use correlations use sample size-based weights that are independent of observation-level variation in correlations. Column (4) reports how many meta-analyses have an effect size that is a correlation and also test for publication bias. Column (5) reports how often these meta-analyses used an Egger's type test (E), a Begg's rank correlation test (B), or a funnel plot (F) to test for publication bias. Column (6) does the same as Column (5) but only reports if the meta-analyses used an Egger's type test (E) or Begg's rank correlation test (B).

TABLE 14
Correcting for Publication Bias

Discipline	Percent Testing (1)	Percent Finding (2)	Percent Correcting(1) (3)	Percent Correcting(0) (4)
Anatomy & Physiology	81%	37%	40%	19%
Biology	75%	35%	42%	8%
Business & Economics	74%	43%	75%	18%
Education	85%	29%	52%	27%
Engineering	58%	17%	50%	9%
Environmental Sciences	58%	24%	29%	6%
Medicine	61%	33%	35%	11%
Pharmacy, Therapeutics & Pharmacology	74%	16%	0%	6%
Psychology	90%	39%	63%	37%
Public Health	69%	35%	42%	12%
Overall	**73%**	**31%**	**47%**	**14%**

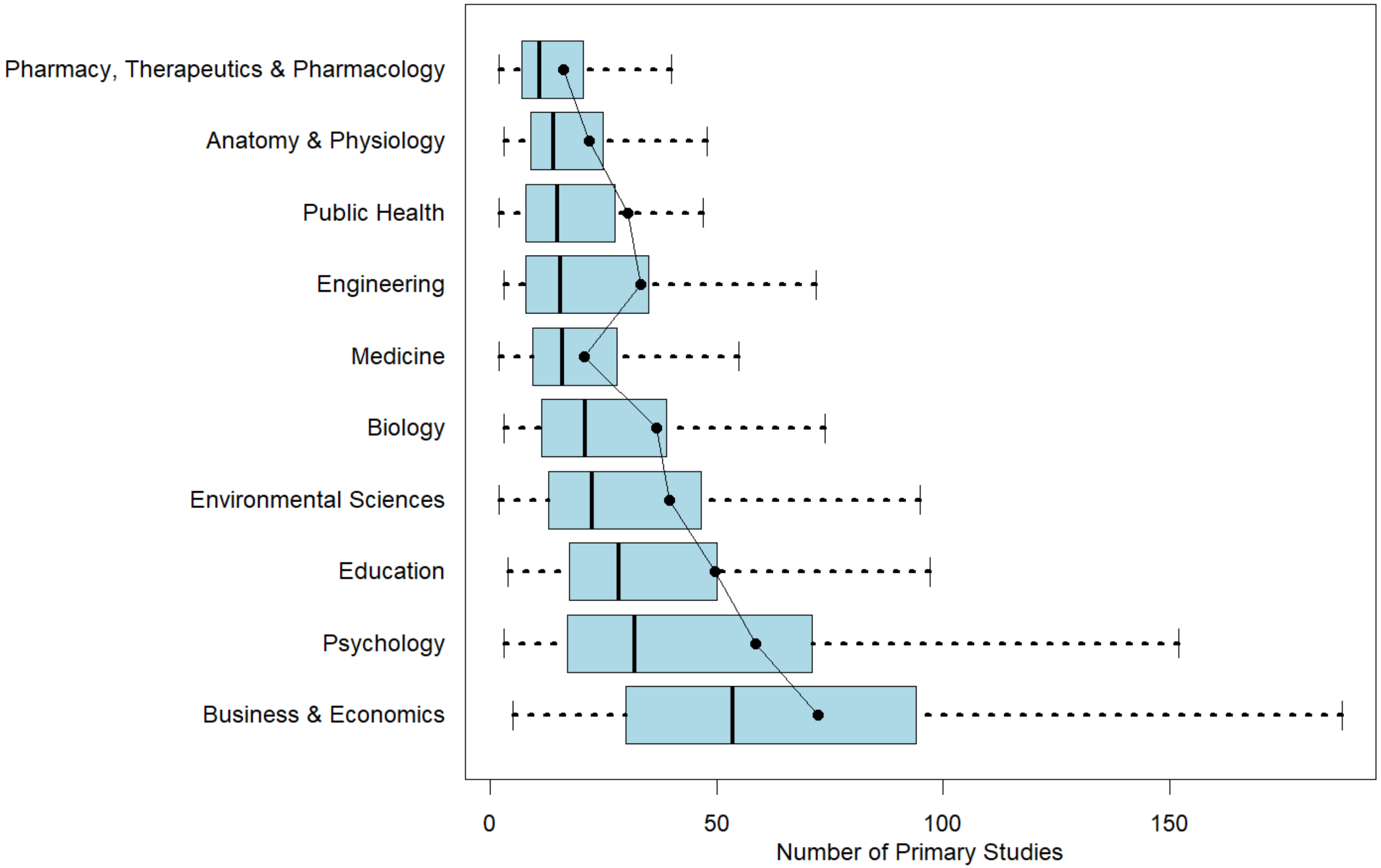
NOTE: Columns (1) and (2) of the table reproduce Columns (1) and (2) from TABLE 8. Column (3) reports the percent of meta-analyses finding publication bias that then went on to correct for it. Column (4) reports the percent of meta-analyses that didn't find publication bias that still went on to correct for it.

TABLE 15
Meta-Analyses that Account for Dependencies

Discipline	No. with Dependencies (1)	% (No.) Addressing (2)	Of (2), % Select (3)	Of (2), % Avg (4)	Of (2) , % MLM (5)	Of (2), % MVMA (6)	Of (2, % CR1 (7)	Of (2), % CR2 (8)
Anatomy & Physiology	46	43% *(20)*	20%	35%	25%	0%	5%	25%
Biology	48	44% *(21)*	19%	14%	76%	10%	5%	0%
Business & Economics	84	73% *(61)*	7%	21%	39%	7%	39%	7%
Education	69	67% *(46)*	7%	35%	24%	2%	2%	48%
Engineering	50	22% *(11)*	36%	45%	9%	9%	0%	18%
Environmental Sciences	78	40% *(31)*	42%	23%	52%	0%	0%	0%
Medicine	33	15% *(5)*	20%	0%	80%	0%	0%	0%
Pharmacy, Therapeutics & Pharmacology	39	10% *(4)*	50%	25%	25%	0%	0%	0%
Psychology	77	77% *(59)*	31%	42%	34%	0%	2%	15%
Public Health	47	23% *(11)*	27%	36%	36%	0%	9%	0%
Overall	**571**	**47%** ***(269)***	**21%**	**30%**	**38%**	**3%**	**11%**	**16%**

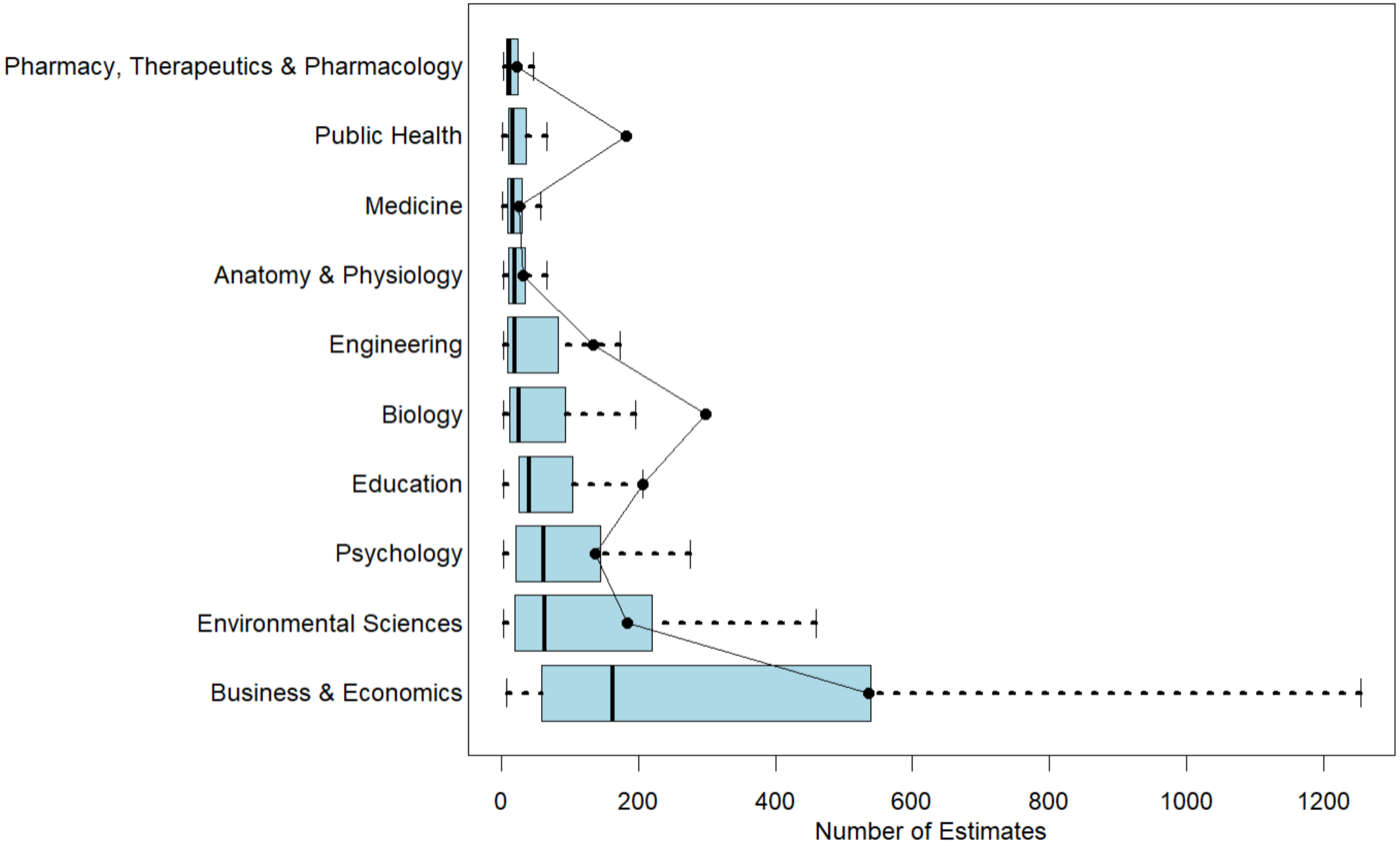
NOTE: Column (1) reports the number of meta-analyses in each discipline that have more estimates than studies, generating a dependency between estimates. The top number in Column (2) reports the percent of meta-analyses that have a dependency that address it using one or more of the following approaches: (i) selecting a subset of estimates from the primary studies, (ii) averaging estimates from the primary studies, (iii) using a multilevel estimator (MLM), (iv) using a multivariate estimator (MVMA), (v) using a CRVE1 clustered robust estimator of the standard error, and (vi) using a CRVE2 estimator. The italicized number below the top number in Column (2) is the number of meta-analyses that addressed dependency. Columns (3) through (8) report conditional probabilities. That is, conditional on addressing dependency, what percent of meta-analyses used that particular approach (e.g., selection, averaging, etc.). For example, 46 meta-analyses in Anatomy & Physiology have more estimates than studies. 43% of these (or 20 of the 46 meta-analyses), address this in some way. Of those 20 meta-analyses, 20% selected a subset of the estimates from the primary studies, 35% averaged some estimates from the primary studies, 25% used a multilevel estimator such as the three-level estimator to estimate the model, no studies used a muiltivariate estimator, 5% adjusted standard errors using the CRVE1 estimator, and 25% used the CRVE2 estimator. Notice that the sum of individual percentages is greater than 100% because some meta-analyses used more than one of these methods to address dependency in their datasets.

FIGURE 1
Number of Studies in a Meta-Analysis by Discipline



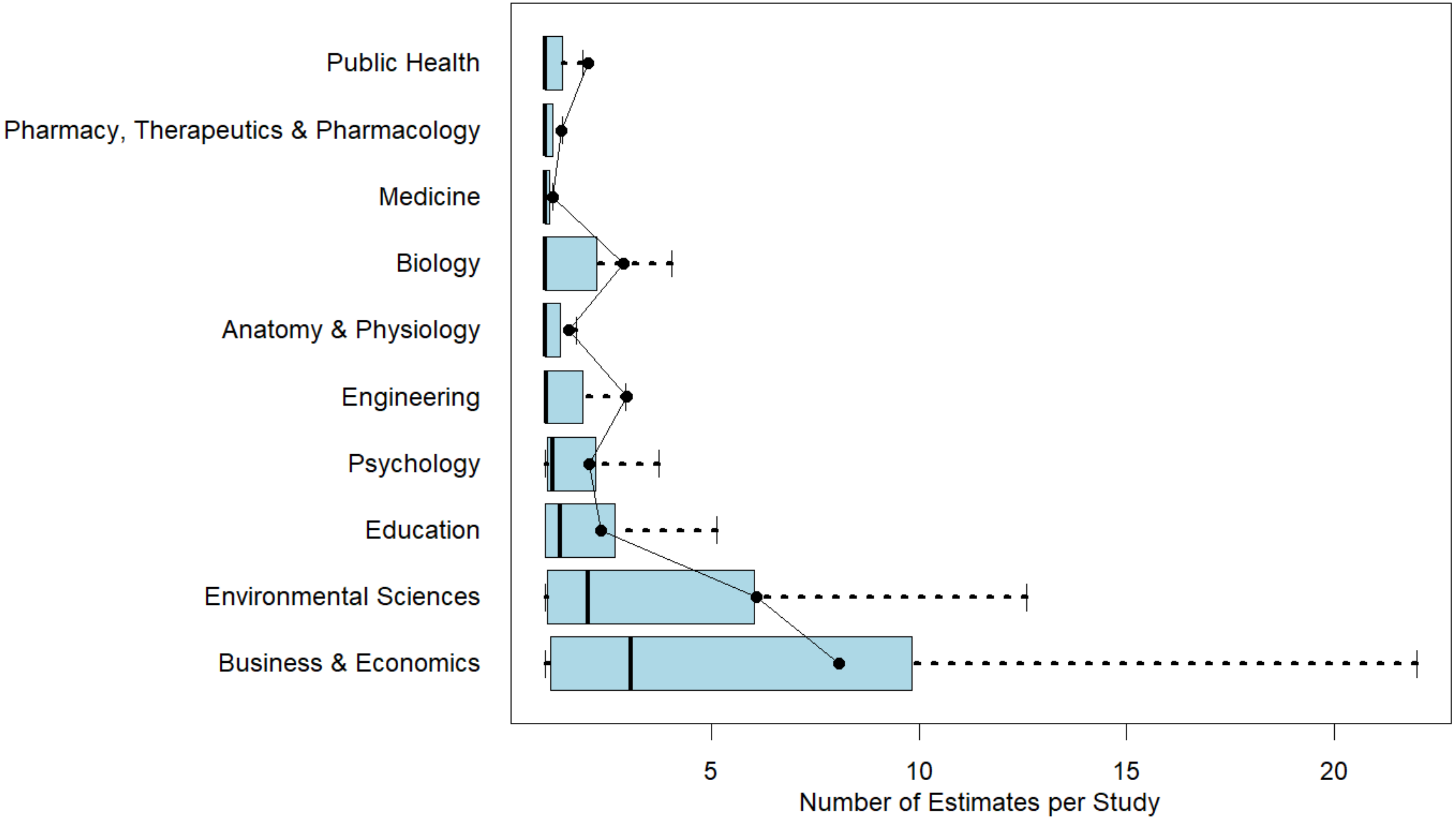
<u>NOTE</u>: FIGURE 1 reports boxplots (without outliers) for each of the ten disciplines with respect to number of studies. Superimposed on the boxplots are the average number of studies per meta-analysis. The figure is arranged with smallest median number of studies per meta-analysis at the top of the figure to largest median number of studies at the bottom of the figure.

FIGURE 2
Number of Estimates in a Meta-Analysis by Discipline



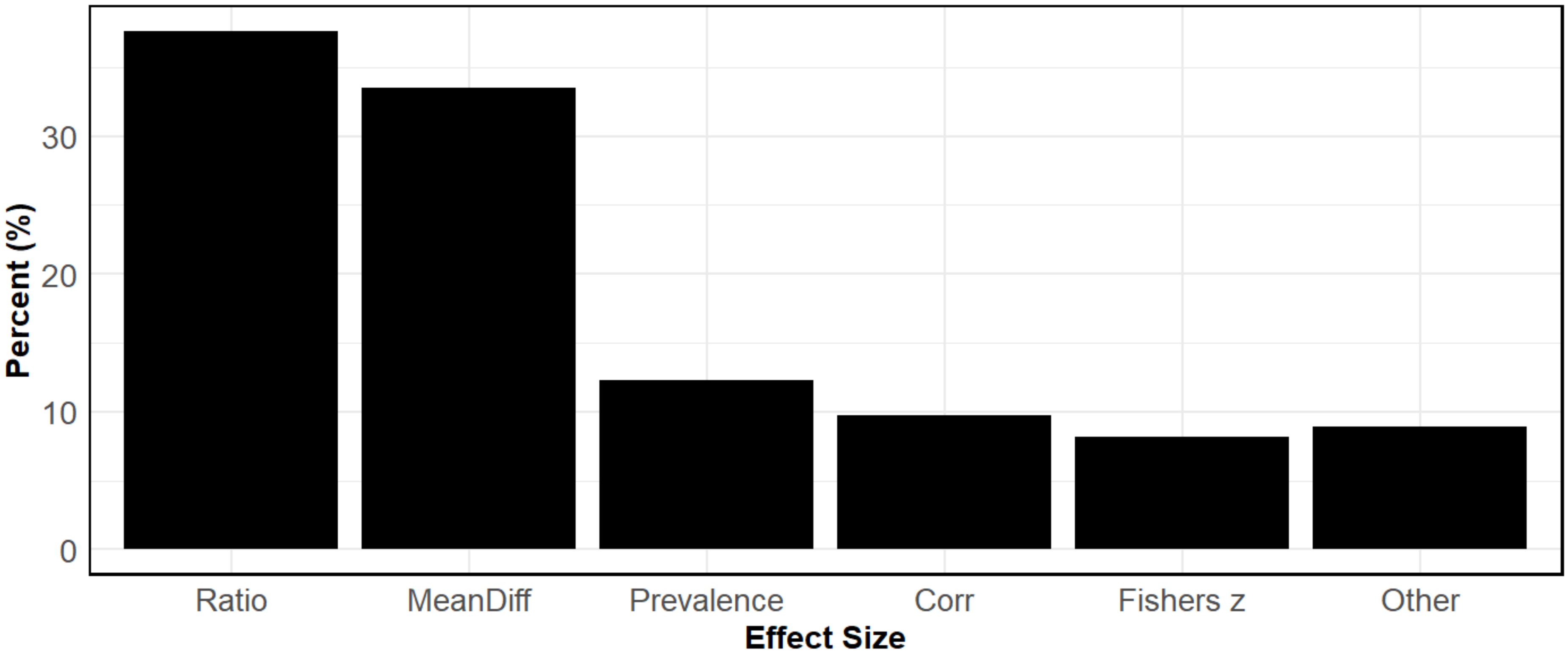
NOTE: FIGURE 2 reports boxplots (without outliers) for each of the ten disciplines with respect to number of estimates. Superimposed on the boxplots are the average number of studies per meta-analysis. The figure is arranged with smallest median number of estimates per meta-analysis at the top of the figure to largest median number of estimates at the bottom of the figure.

FIGURE 3
Number of Estimates per Study in a Meta-Analysis by Discipline



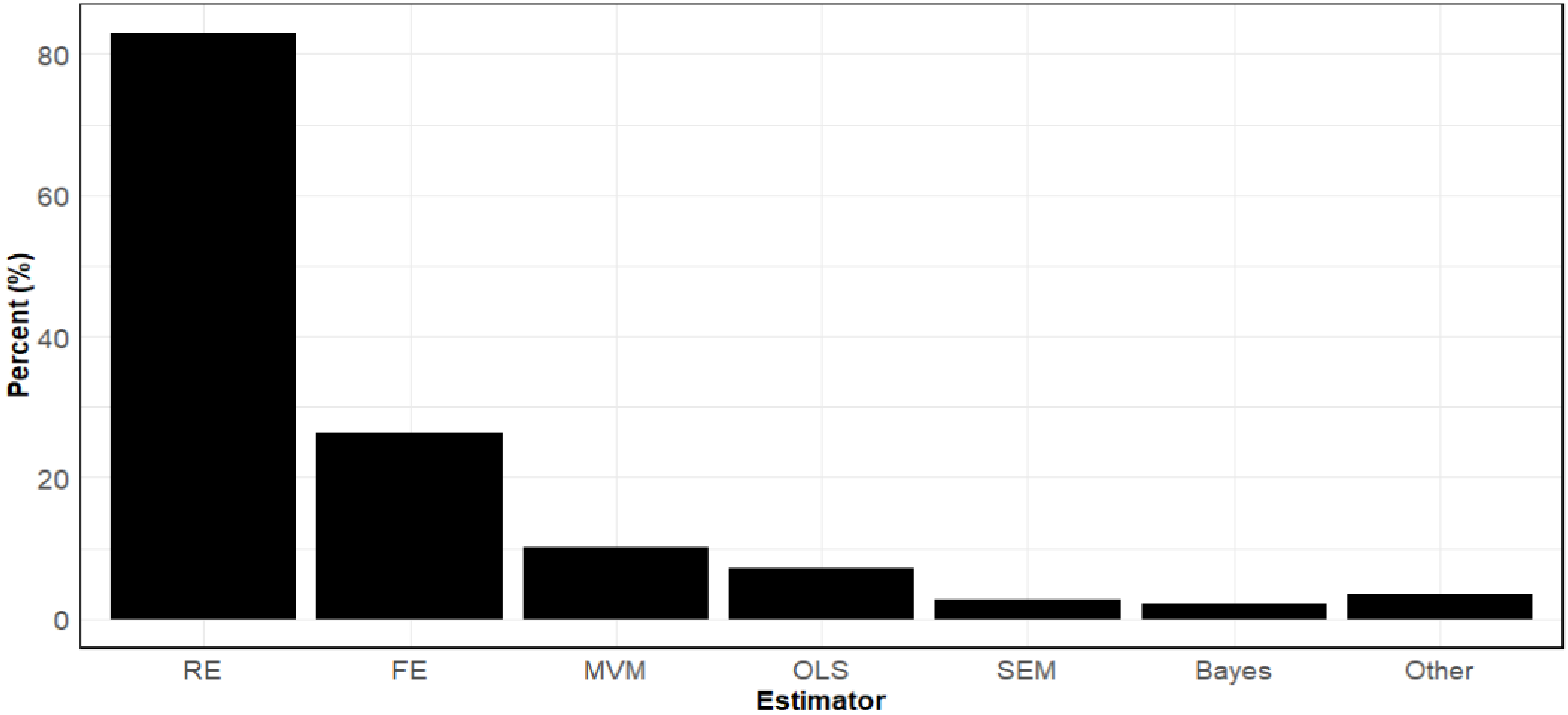
<u>NOTE</u>: FIGURE 3 reports boxplots (without outliers) for each of the ten disciplines with respect to number of estimates per study. Superimposed on the boxplots are the average number of estimates per study. The figure is arranged with smallest median number of estimates per study at the top of the figure to largest median number of estimates per study at the bottom of the figure.

FIGURE 4
Usage Rates of Different Effect Sizes



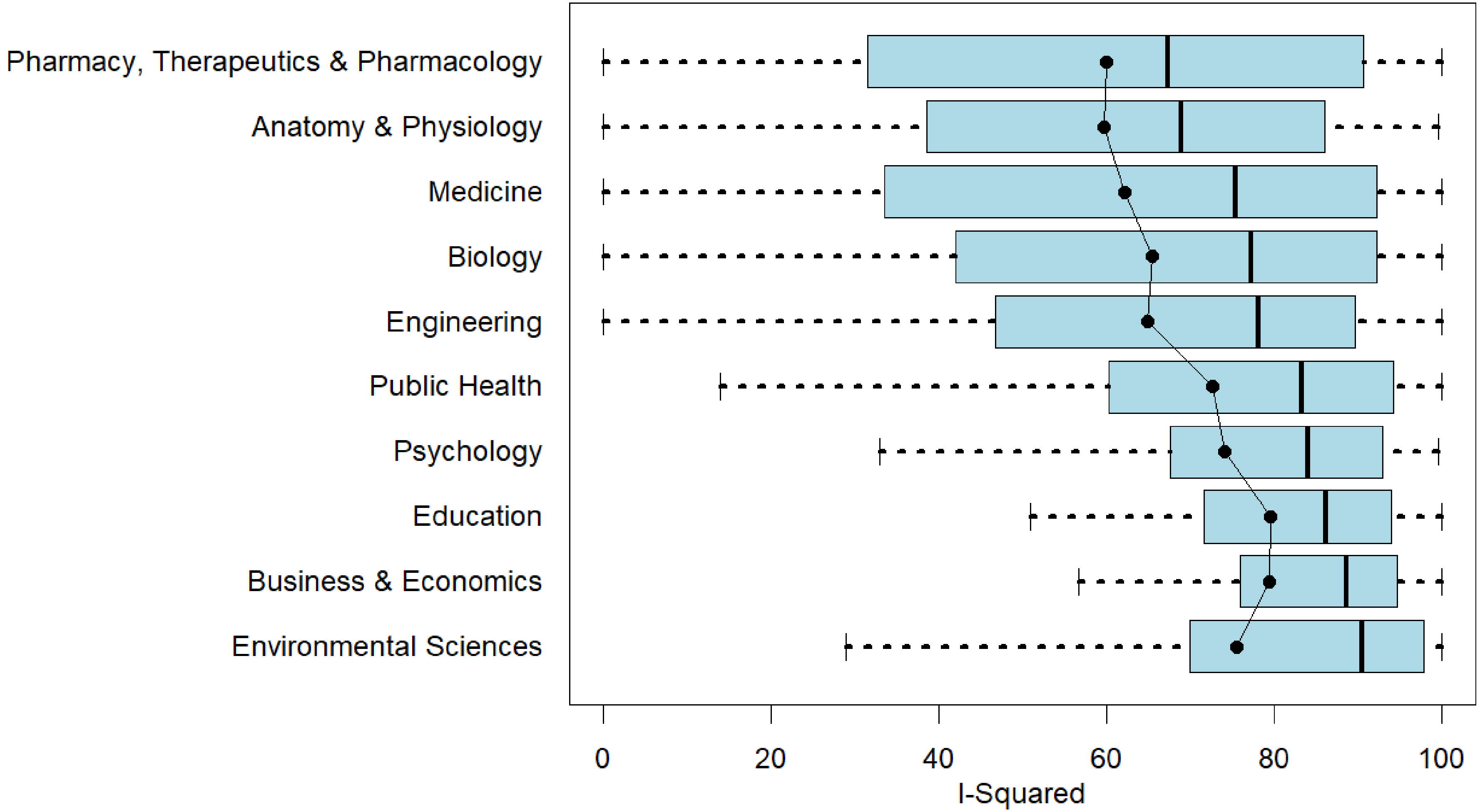
NOTE: FIGURE 4 reports aggregate usage rates of different effect sizes across all disciplines. "Ratio" includes odds ratios, hazard ratios, and risk ratios. "Mean Diff" includes Cohen's d and Hedge's g. "Prevalence" uses a frequency or count to measure outcome. "Corr" stands for correlation and includes partial correlation coefficients.

FIGURE 5
Usage Rates of Different Meta-Analytic Estimators



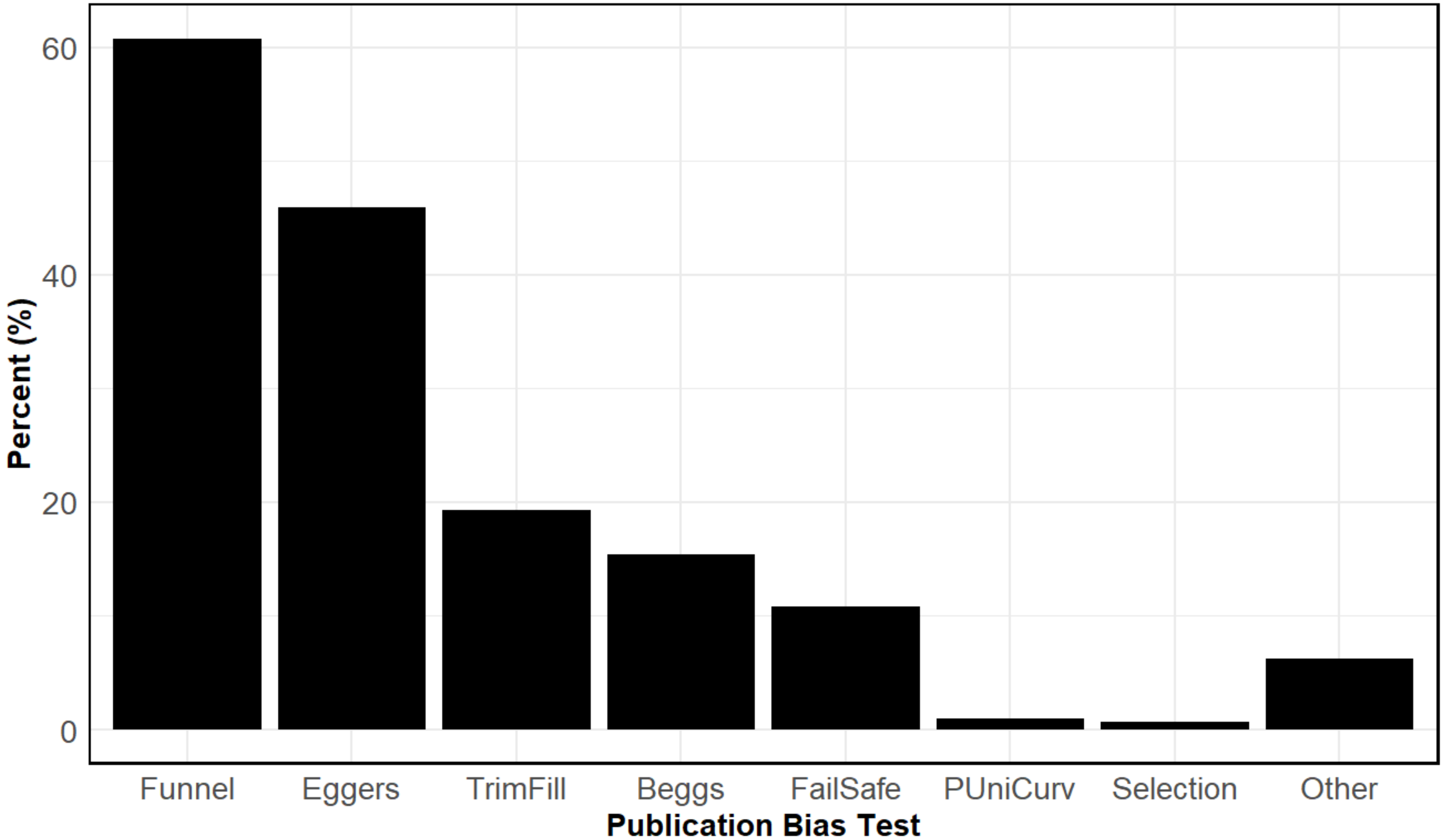
<u>NOTE</u>: FIGURE 5 reports aggregate usage rates of different meta-analytic estimators across all disciplines. RE = Random Effects; FE = Fixed Effects; MVM = Multi-Level Model; OLS = Ordinary Least Squares; SEM = Structural Equation Modelling; Bayes = Bayesian estimation.

FIGURE 6
I-Squared by Discipline



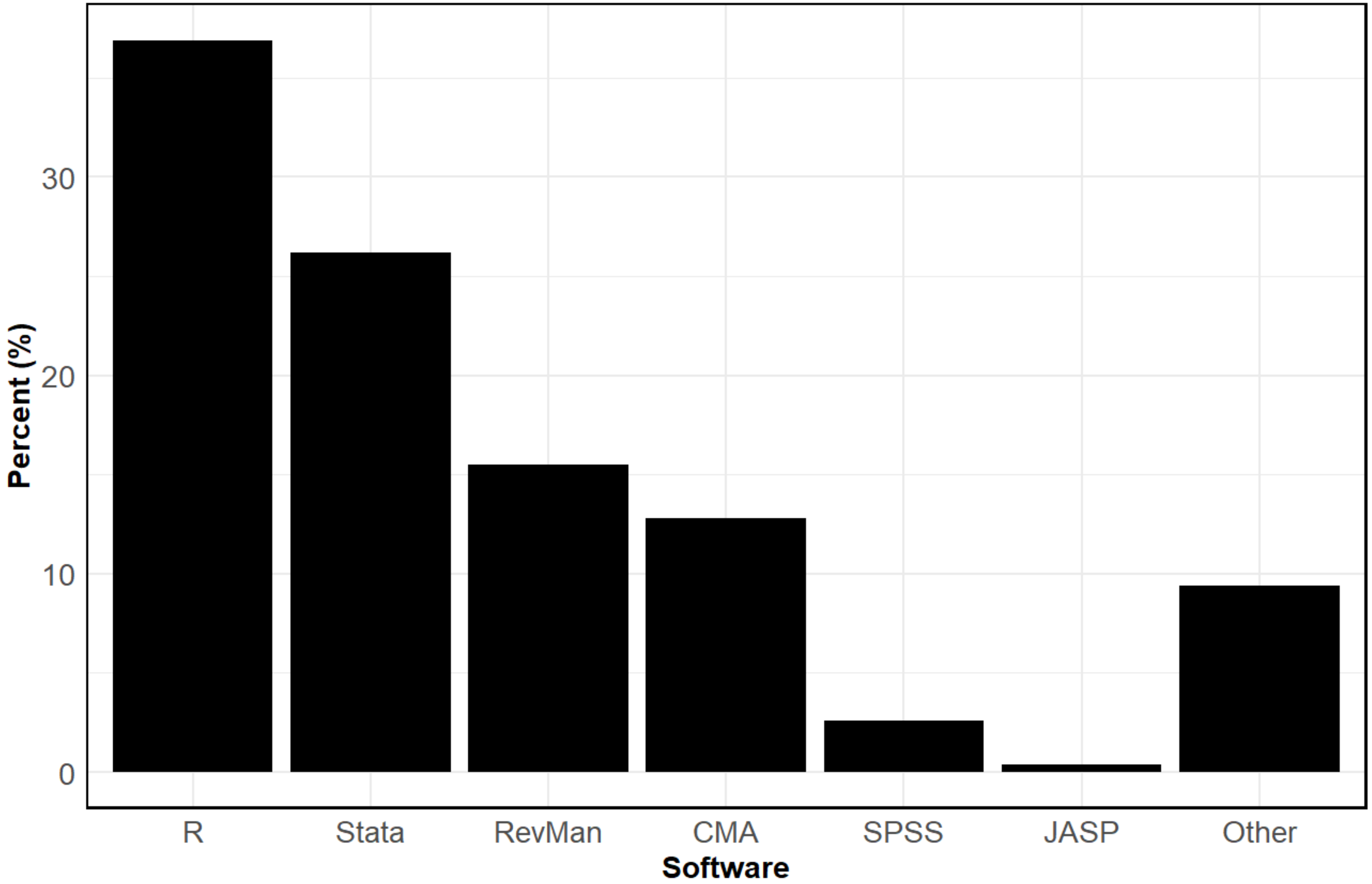
NOTE: FIGURE 6 reports boxplots (without outliers) for each of the ten disciplines with respect to I-squared as a relative measure of effect heterogeneity. Superimposed on the boxplots is the mean I-squared value for that discipline. The figure is arranged with smallest median I-squared value at the top of the figure to the largest median I-squared value at the bottom of the figure.

FIGURE 7
Prevalence of Different Types of Tests for Publication Bias



NOTE: FIGURE 7 reports aggregate usage rates of different types of publication bias tests. Funnel = Funnel plot; Eggers = Egger-type regression; TrimFill = Trim and Fill; Beggs = Begg and Mazumdar's rank correlation test; FailSafe = Fail Safe N; Selection = publication bias uses a selection model; PUniCurv = either p-Uniform or p-Curve test for publication bias.

FIGURE 8
Usage Rates of Different Statistical Packages



NOTE: FIGURE 8 reports aggregate usage rates of different statistical software packages.