INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

REDUCING THE BIASES OF THE CONVENTIONAL META-ANALYSIS OF CORRELATIONS

T. D. Stanley
Hristos Doucouliagos
Tomas Havranek

IES Working Paper 34/2023

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Stanley, T. D., Doucouliagos H., Havranek T. (2023): " Reducing the Biases of the Conventional Meta-Analysis of Correlations" IES Working Papers 34/2023. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Reducing the Biases of the Conventional Meta-Analysis of Correlations

T. D. Stanley[1]
Hristos Doucouliagos[2]
Tomas Havranek[3]

[1]Department of Economics, Deakin University, Victoria, Australia & Meta-Research Innovation Center at Stanford. E-mail: tom.stanley1@deakin.edu.au
[2]Department of Economics, Deakin University, Victoria, Australia.
E-mail: douc@deakin.edu.au
[3]Charles University, Prague, Centre for Economic Policy Research, London, and Meta-Research Innovation Center at Stanford. E-mail: tomas.havranek@fsv.cuni.cz

December 2023

Abstract:
Conventional meta-analyses of correlations are biased due to the correlation between the estimated correlation and its standard error. Simulations that are closely calibrated to match actual research conditions widely seen across correlational studies in psychology corroborate these biases and suggest a solution. UWLS+3 is a simple inverse-variance weighted average (the unrestricted weighted least squares) that adjusts the degrees of freedom and thereby reduces small-sample bias to scientific negligibility. UWLS+3 is also less biased than conventional random-effects estimates of correlations and Fisher's z, whether or not there is publication selection bias. However, publication selection bias remains a ubiquitous source of bias and false positive findings. Despite the correlation between the estimated correlation and its standard error even in the absence of any selective reporting, the precision-effect test/precision-effect estimate with standard error (PET-PEESE) nearly eradicates publication selection bias. PET-PEESE keeps the rate of false positives (i.e., type I errors) within their nominal levels under the typical conditions widely seen across psychological research and with or without publication selection bias.

JEL: C83
Keywords: correlations, meta-analysis, publication selection bias, small-sample bias

Acknowledgement: Havranek acknowledges support from the Czech Science Foundation (#21-09231S) and from NPO "Systemic Risk Institute" LX22NPO5101, funded by the European Union - Next Generation EU (Czech Ministry of Education, Youth and Sports, NPO: EXCELES). An online appendix is available at meta-analysis.cz/correlations.

1. Introduction

Correlations are widely used to summarize psychological research via inverse-variance weighted meta-analysis in spite of the fact that, by conventional definitions, the variance (and standard error, SE) of correlations is a function of the correlation estimate itself. What has yet to be fully recognized is that this dependence of the variance on the size of the correlation causes all conventional meta-analysis of correlations to be biased (Stanley, Doucouliagos, Maier et al., 2023). The conventional approach to this dependence is to employ Fisher's *z* transformation, as its SE is independent of the estimate of *z* (e.g., Borenstein *et al*., 2009).[1] Yet, many meta-analyses of simple, untransformed correlations are routinely conducted in psychology. For example, a survey found that the majority meta-analysis published in the *Psychological Bulletin* (108 of 200) concerned correlations. Within these 108 meta-analyses of correlations (84.3%) did *not* use the Fisher's *z* transformation, but rather, the simple untransformed correlations (Stanley *et al*., 2018).

We follow previous studies that found that conventional meta-analyses of bivariate and partial correlations and are biased, largely due to estimated correlations being inversely correlated with their variances (Stanely and Doucouliagos, 2023; Stanley, Doucouliagos, Maier et al., 2023). Fortunately, these biases are small-sample biases. A new estimator, UWLS_{+3}, is introduced below that reduces these biases to scientific negligibility by making a simple adjustment to the degrees of freedom. However, these studies assumed that the sample sizes were constant across all studies within a meta-analysis and that there was no publication selection bias (PSB). While these assumptions are necessary to isolate and to identify the small-sample bias caused by correlation's mechanical inverse correlation with its own SE, these conditions do not hold, even approximately, for most meta-analyses of social science research.

First, the range of sample sizes synthesized by the typical meta-analysis is many times its median value. Thus, at least some studies in most meta-analyses will be sufficiently large to reduce correlation's small-sample bias to practical negligibility. Second, although not every area of research selects for statistical significance and thereby produces PSB, it is rare when PSB can be ruled out *a priori*. When present, PSB can be substantial, creating high rates of false positives in conventional meta-analyses (Kvarven, Strømland and Johannesson, 2020); also see Tables 3

[1] It should be noted that Fisher's *z* transformation is not the same as the z-values, widely used throughout statistics and meta-analysis to represent the normal distribution.

below. In this study, we show that a small-sample correction, $UWLS_{+3,}$ reduces bias to negligibility and investigate whether conventional meta-analysis will still be biased when there is a wide range of sample sizes and heterogeneity, with and without accompanying selection for statistical significance. In short, conventional, inverse-variance weighted meta-analyses are still biased under typical research conditions seen in psychology. However, we do not stop there. We also identify those meta-analyses methods that have no notable biases with or without PBS as well as those that are able to maintain their nominal type I errors (that is, those do not have inflated rates of false positives) even with publication bias.

2. Correlation and its Variances

The conventional formula for the Pearson (bivariate) correlation coefficient, *r*, is:

$$r = \sum(X_i - \bar{X})(Y_i - \bar{Y}) \Big/ \left(\sqrt{(X_i - \bar{X})^2} \cdot \sqrt{(Y_i - \bar{Y})^2}\right) \qquad i = 1, 2, \ldots, n. \qquad (1)$$

When testing whether the association between X and Y is statistically significant, *r*'s variance is:

$$S_1^2 = (1 - r^2) \Big/ (n - 2) \qquad (2)$$

$t = \frac{r}{S_1}$ is the conventional test statistic for testing the hypothesis: H_0: $\rho = 0$. Correlation's *t*-value is also equal to the *t*-value for the slope coefficient from the simple linear bivariate regression between *X* and *Y*. See Stanley, Doucouliagos, Maier et al. (2023) Supplement for a numerical analysis proof.

In contrast, conventional meta-analysis uses a different variance for correlations:

$$S_2^2 = (1 - r^2)^2 \Big/ (n - 1) \qquad (3)$$

(Hunter and Schmidt, 1990; Borenstein et al., 2009, among many others). Note that the differences between these variance formulae are: S_2^2 squares $S_1^2{}'s$ numerator, $1 - r^2$, and $S_2^2{}'s$ degrees of freedom, $(n-1)$, are one fewer. Because -1≤ $r \leq 1$ and $(n-1) > (n-2)$, $S_2^2 < S_1^2$ for all sample sizes and $|r| \neq \{0 \text{ or } 1\}$. Simulations reported in Table 1, below, Table S1 of the Supplement, and in Stanley, Doucouliagos, Maier et al., (2023) establish that using S_2^2 causes conventional meta-analyses to be twice as biased as those which use S_1^2 and the CI's produced by S_2^2 systematically cover less than 95% of the estimated mean correlations—also see Table 1, below.

Lastly, there are different ways to calculate correlations. Following Gustafson (1961) and Fisher (1921), Stanley, Doucouliagos, Maier et al. (2023) demonstrate that eq. (4), below, gives the exact same values for estimated correlations as the more conventional correlation formula, Eq (1).

$$r = {}^{t}\Big/_{\sqrt{t^2 + df}} \tag{4}$$

Where $df = n - 2$. t is the conventional t-test for the statistical significance of the slope coefficient of a bivariate regression or, equivalently, of the t-value of correlations using S_1 . This t-formula for correlations, eq. (4), is central for our new small-sample correction, UWLS_{+3}.

3. Conventional Meta-analysis of Correlations

Random-effects (RE) weighted averages are, by far the most, commonly employed meta-analysis approach used to systematically review and summarize correlations across studies in a given area of research in psychology. RE is, thereby, the conventional standard upon which to establish the bias of the conventional meta-analyses of correlations—see Table 1, below. RE serves as the baseline from which to evaluate the statistical performance of alternative meta-analysis methods.

3.1 The Unrestricted Weighted Least Squares Weighted Average

The unrestricted weighted least squares (UWLS) is an alternative simple weighted average that has statistical properties practically equivalent to RE under ideal conditions for RE and notably superior if there is publication bias or if small-sample studies are more heterogeneous (Stanley and Doucouliagos, 2015 & 2017, Stanley, Doucouliagos and Ioannidis, 2017 & 2022b). Also, UWLS has been shown to be widely and notably superior to RE in most applications in psychology and medicine (Stanley, Doucouliagos and Ioannidis, 2022b; Stanley, Ioannidis, Maier *et al*., 2023).

UWLS is calculated from the simple meta-regression:

$$t_j = \frac{r_j}{SE_j} = \alpha_1 \left(\frac{1}{SE_j}\right) + u_j \qquad j = 1, 2, \ldots, k \qquad (5)$$

Where SE_j is SE of the j[th] correlation calculated as the square root of either S_1^2 or S_2^2 from their respective formulas; that is, eq. (2) or eq. (3), above. Any standard statistical software for regression analysis will automatically estimate UWLS (the slope coefficient, $\hat{\alpha}_1$), its standard error and, *p*-value.[2]

Stanley, Doucouliagos and Havránek (2023) offer a new correction, $UWLS_{+3}$, for the small-sample biases of the conventional meta-analysis of partial correlations first identified in Stanley and Doucouliagos (2023). Table S1 of Appendix A of the supplement reports the biases of the conventional meta-analysis of *bivariate* correlations. Like the biases of the meta-analysis of partial correlations, they are positive and can be of a notable magnitude for small samples. Following Fisher (1921), "sampling distribution of the partial correlation obtained from *n* pairs of values, when one variable is eliminated, is the same as the random sampling distribution of a total correlation derived from (*n*-1) pairs. By mere repetition of the above reasoning it appears that when *s* variates are eliminated the effective size of the sample is diminished to (*n*-*s*)" (p. 330). We have verified Fisher's insightful observation also works for partial correlations and vice versa. For both correlations and partial correlations, we have found that the small-sample biases of conventional

[2] For example, a simple R program is:

```
#read in your data file that has a column for
# correlations, labeled "r", and standard errors
# as the sqrt of eq(2) labeled "ser".
pathName = "C:/Users/USER/Documents/MyData.csv"
dat = read.csv(pathName)
r = dat$r
ser = dat$ser
k = length(r) #number of studies
t = r/ser
Precision=1/ser
reg = lm(t ~ 0 + Precision)
UWLS = as.numeric(reg$coefficients)
```

meta-analysis is nearly a perfect function of the inverse degrees of freedom (adjR^2 = .999985)—see Appendix A of the supplement. Thus, simple adjustments to the degrees of freedom, $UWLS_{+3}$, can reduce these small-sample biases to scientific triviality. See Appendix A of the supplement for the numerical analysis of the small-sample biases of the meta-analysis of bivariate correlations.

$UWLS_{+3}$ is the unrestricted weighted least squares weighted average (i.e., eq. (5)) after three is added to the degrees of freedom in eq. (4), giving:

$$r_3 = t/\sqrt{t^2 + n + 1} \tag{6}$$

These *t*-values are the *t*-values from the estimated bivariant regression slope coefficient or they may be equivalently calculated from the conventional *t*-value for correlations, $t = \frac{r}{S_1}$. Again, correlations calculated from eq. (4), with $df = n - 2$, produce identical correlation values as those calculated from conventional formulas for correlations. We focus on $UWLS_{+3}$ rather than corrections to RE because adjustments to UWLS have notably smaller biases than RE and RE's small-sample correction in the presence of publication selection bias (PSB).

3.2 Fisher's z Transformation

An issue that has long been recognized by meta-analysts is that the SEs of correlations are a mathematical function of the correlation itself—recall eqs. (2) and (3). This dependence is the source of the small-sample bias of the meta-analysis of correlations (Stanley, Doucouliagos, Maier et al., 2023). Strictly speaking, the inverse-variance weights are no longer optimal and create a bias. To circumvent this issue, meta-analysts often first transform them to Fisher's *z*, calculate the random-effects estimate, then convert this RE estimate from terms of *z* back to a correlation (Bornstein *et al*., 2009).[3] Here, we call this *z*-transformed RE estimator, REz. Table 2 compares the statistical properties of REz to $UWLS_{+3}$, PET-PEESE, and RE (from Table 1).

[3] Correlations are converted to *z* by: $0.5 \cdot \ln\left[(1+r)/(1-r)\right]$, and Fisher's *z* is transformed back to correlations by: $(e^{(2z)} - 1)/(e^{(2z)} + 1)$.

3.3 PET-PEESE Model of Publication Selection Bias

Publication selection bias (PSB), variously called: the "file drawer problem," "publication bias," "reporting bias," "p-hacking," and "questionable research practices" (QRP), has long been recognized by social scientists and medical researchers as a central problem for meta-analysis. PSB has been offered as a leading explanation of the widely discussed 'replication crisis', and recent meta-research surveys have shown that PSB is the central suspect in the exaggeration of psychology's typical reported effect sizes and statistical significance (Klein, Vianello, Hasselman, *et al*., 2018; Bartoš, Maier, Wagenmakers, *et al*., 2023; Bartoš, Maier, Shanks, *et al*., 2023).

PET-PEESE ranks among the better methods to accommodate and reduce PSB (Bartoš, Maier, Wagenmakers, *et al*., 2023; Carter, Schonbrodt, Gervais & Hilgard, 2019). PET-PEESE is calculated as the slope coefficient from one of two meta-regressions:

$$r_j = \delta_0 + \delta_1 SE_j + u_j \tag{7}$$

$$r_j = \gamma_0 + \gamma_1 SE_j^2 + e_j \tag{8}$$

using weighted least squares (WLS) with $1/SE_j^2$ as the weights (Stanley and Doucouliagos, 2014). If the regression coefficient, δ_0, is statistically significant (one-tail α=.10), then the estimate of γ_0 is PET-PEESE. Otherwise, the estimate of the regression coefficient, δ_0, is PET-PEESE.

PET-PEESE has been used in dozens of meta-analyses in psychology. For example, PET-PEESE anticipated the failure of ego depletion to replicate (Carter *et al*., 2015; Haggard *et al*., 2016). Kvarven, Strømland, and Johannesson (2020) conducted a systematic review of all pairs of preregistered multi-lab replications and meta-analysis. They compared RE, 3PSM (i.e., selection models of publication bias), and PET-PEESE to the findings from these large-scale preregistered, multi-lab replications. On average, RE was three times larger than the corresponding replication result, bias = .26 *d* (Cohen's *d*), and RE had a 100% 'false positive' rate (Kvarven, Strømland, and Johannesson, 2020), and 3PSM was little better. In contrast, PET-PEESE's bias, relative to these preregistered multi-lab replications, is only .051*d*, and PET's false positive rate is much lower than

RE's, especially so (9%) when Cohen's (1990) probabilistic proof of a null effect defines 'false positive' (Stanley, Doucouliagos, and Ioannidis, 2022a).[4]

Incidentally, PET-PEESE belongs to the same family of UWLS estimators as $UWLS_{+3}$ and along with other methods that reduce publication bias: WAAP (weighted average of the adequately powered)and WILS (weighted and iterated least squares) (Stanley, Doucouliagos, and Ioannidis, 2017; Stanley and Doucouliagos, 2022).[5] UWLS may be seen as a PET-PEESE meta-regression model that uses the same weights but does not include any independent variable: neither *SE* nor SE^2.

Table 2, below, also reports simulations of a second version of PET-PEESE that regresses Fisher's *z* on its *SE* or variance. PPz first converts correlations to Fisher's *z*, regresses these *z*s using corresponding versions of Eqs. (7) and (8), and then transforms PPz back to a correlation. PPz avoids the correlation of *r* and its *SE* when there is no publication bias. Another alternative solution, which we do not simulate here, would be to use the square root of the inverse sample size as an instrument for the standard error in PET-PEESE (Stanley, 2005; Irsova, Bom, Havranek, and Rachinger, 2023). Among other things, the instrumental-variable PET-PEESE technique accounts for the mechanical correlation between *r* and its SE. See Irsova, Bom, Havranek, and Rachinger (2023) for simulations of the instrumental estimator.

3.4 An Illustration

Eastwick *et al*. conducted a meta-analysis of the correlations of physical attractiveness and earning potential on men and women's romantic evaluations (Eastwick, Luchies & Finkel *et al*., 2013). The research literature suggests that: "The attractiveness of the target affects men's romantic evaluations more than women's, and the earning prospects of the target affect women's romantic evaluations more than men's" (Eastwick, Luchies & Finkel *et al*., 2013, p. 627). This meta-analysis reported several random-effects estimates but focused on the gender differences and their moderators. For the sake of illustrating the methods discussed above, we focus on the correlation between the perceived earning potential of candidate men on women's romantic evaluations.

---

[4] Only the precision-effect test (PET) provides a valid test (H_0: δ_1=0, from Eq. (7) for the presence of a nonzero mean effect, after correcting for potential PSB.

[5] Both WAAP and WILS calculate UWLS on subset of the effect sizes. WAAP uses only those studies that have 80% or higher power while WILS first removes those estimates most responsible for excess statistical significance.

Conventional random effects estimate the correlation of the earnings potential of the target on women's romantic evaluations as 0.128 – 95% CI (0.092, 0.164), k=73. Using REz does not notably reduce these values– 0.127, 95% CI (0.092, 0.163). This is to be expected as the correlation is small, and small correlations have smaller small-sample biases. Furthermore, the sample sizes vary widely from 11 to over 7,000 with most studies having n > 100.

On the other hand, $UWLS_{+3}$ reduces the RE estimate by over 60%, 0.050, 95% CI (0.022, 0.078). That is, $UWLS_{+3}$ reduces a small correlation to a trivial one by Cohen's benchmarks. It is important to note that Eastwich *et al.* (2013) accept Cohen's definition of 'small' effect sizes ($.1 \leq r \leq .3$) and use it to characterize their central findings—(Eastwick *et al.*, 2013, Abstract). $UWLS_{+3}$ is calculated by first adjusting each correlation by eq. (6), giving r_3, then applying the simple UWLS regression, eq. (5), of t-values = r_3/S_1 (DV) with precision, $1/S_1$, as the only independent variable and no constant—see eqs (6) and (5) and the Supplement for the STATA code.

The primary reason that $UWLS_{+3}$ notably reduces the effect size is likely publication selection bias. UWLS, in general, is widely known to reduce PSB more than corresponding random-effects, and the below simulations confirm that $UWLS_{+3}$ is less biased than RE and REz. However, these simulations also show that all weighted averages are notably biased when there is a small correlation, PSB, and notable heterogeneity, as we see here (tau = 0.128; I^2 = 85%).

Testing whether the coefficient on SE in eq. (7) is statistically significant is a test for PSB (the Egger test), also called the funnel-asymmetry test, or FAT (Egger Smith & Schneider & Minder, 1997; Stanley, 2005; Stanley and Doucouliagos, 2014). The estimated FAT-PET meta-regression, eq. (7), for these earnings-romance correlations and the associated Fisher's z (Fz) are:

$$r_j = -.026 + 1.94 \cdot SE_j \qquad (9)$$
$$t = (-1.37) \quad (5.19)$$

$$Fz_j = -.029 + 1.94 \cdot SE_j \qquad (10)$$
$$t = (-1.55) \quad (5.39)$$

Where the second line reports the t-values of the intercept (PET) and slope coefficients (FAT) in parentheses and both meta-regressions use inverse variances as WLS weights. Also note that SE_j is different in eqs. (9) and (10). For correlations, S_1 is its standard error, and Fisher's z employs 1/sqrt(n-3) as its standard error. In both cases, PET fails to reject the null hypothesis that earnings-romance correlation for women is zero (t= {-1.37; -1.55}; p > .05). In other words, once potential publication selection bias (or small-study bias or funnel asymmetry) is accommodated, no evidence of a positive earnings-romance correlation remains.[6] Also, both tests of the slope coefficients (FAT) are consistent with funnel asymmetry and, therefore, PSB (t= {5.17; 5.39}; p < .001).[7] This funnel asymmetry is also seen in the funnel graph—see Figure 1.

Consistent with this interpretation, observe that the largest sample estimates are all quite small. For example, there are only two studies that are adequately powered (power $\geq$ 80%), using $UWLS_{+3}$ as the estimate of the population mean. These two studies are at the top of the funnel (Figure 1)— r = {-0.06, 0}. Considerations of power alone make the random-effects estimate dubious, whether Fisher's z is used or not. Greater resilience to PBS is perhaps $UWLS_{+3}$'s most important property. We turn next to simulations that show this to be a general property of $UWLS_{+3}$.

4. Simulations

To better understand the statistical properties of the meta-analysis of correlations under research conditions commonly seen in psychology, we conduct Monte Carlo simulations. Unlike replications or other empirical analyses, simulations allow us to set and thereby know the exact 'true' (population) value, ρ, of the correlations investigated. To ensure that they reflect typical research conditions found across psychology, we closely calibrate our simulations design to match the key research dimensions found in correlational research. For this purpose, we employ 108 *Psychological Bulletin* meta-analyses of correlations reported in Stanley *et al*. (2018). These 108 meta-analyses jointly contain 5,891 pairs of estimated correlations and their standard errors, from which we can also calculate the sample sizes.

---

[6] Because these intercepts are in the opposite direction of the meta-analysis estimates (i.e., negative), we interpret them as negligible. When the PET estimate is of the opposite sign as UWLS, PEESE, eq. (8), should not be calculated. In these cases, there is such a strong correlation with the standard error that any statistical evidence of a mean effect is erased once potential PSB in accommodated. PEESE should be employed only if there is some evidence of an effect in the predominant direction.

[7] Note that the magnitude of the estimated FAT coefficients, 1.94, is quite substantial. When, the FAT coefficient is two or larger, Doucouliagos and Stanley (2013) categorize this as 'severe' publication selection because it implies that the average effect size is exaggerated by twice its SEs; just sufficient to can make a null effect appear statistically significant. The SE of Fisher's z does not depend on the magnitude of the correlation (or Fz); thus, this clear positive correlation with SE cannot be dismissed as a statistical artifact of its variance formula. Nor can the fact that the formula for S_1^2 depends on r be used to dismiss its *positive* correlation, eq. (9), as this formula embeds a slight *negative* correlation.

The median sample size is 95, which we round to 100, the 10th percentile uses a sample of 30, and 90th percentile is 424, which we round down to 400 so as not to exaggerate the likely precision of some studies in an area of research. Although a very small percentage of studies use thousands and tens of thousands of observations, to assume larger sample sizes risks underestimating the biases of the majority of meta-analysis of correlations. Recall that the Supplement (Table S1) demonstrates how all meta-analyses of correlations and partial correlations have small-sample biases which predictably disappear with larger sample sizes at a rate proportional to 1/n. Using these percentiles as our anchors, we fill in the remainder of the sample size distribution, n = {30, 40, 50, 75, 100, 100, 125, 160, 200, 400}, to correspond to the sample size distribution observed across these 108 *Psychological Bulletin* meta-analyses.

Similarly, the values of the population correlation are set to correspond to the observed distribution of random-effects estimates reported in these same 108 meta-analyses. The median absolute value of these 108 REs is 0.232, which, for convenience, we approximate by the sqrt (1/17) = 0.243. The 10th percentile is 0.07, which we 'round' up to sqrt (1/82) = 0.110. As shown in previous studies and we confirm below, small values of ρ produce practically no bias unless study results are selected for their statistical significance (i.e., publication selection bias). Thus, we make this small correlation a bit larger, intentionally. The 90th percentile of the RE distribution is 0.422, which we 'round' up to sqrt (1/5) = 0.447. The 10th and 90th percentiles reflect a range of ρ values likely seen in practice. However, as discussed in Section 3.1, all these values are likely an exaggerated reflection of the 'true' population mean as RE is widely recognized to be highly biased in the presence of PSB, a condition we simulate, corroborate, and discuss further below. Thus, one should focus on the results of the more representative correlation effect size, 0.243 and 0.110, or consider the average across all three values of ρ reported in Tables 1 and 2 as 'representative.'

In the heterogeneity conditions, labelled 'Het' in Tables 1 and 2, we again rely on what was found to be the typical across these 108 meta-analyses, mean I^2 = 64.5%. Note that the typical heterogeneity reported in Tables 1 and 2, below, for the 'Het' case nearly reproduces this level of relative heterogeneity. To do so, we assume that heterogeneity is weakly and inversely correlated with sample size; that is, normally distributed with standard deviations of τ ={.45, .45, .3, .3, .3, .3, .3, .3, .075, .075} as n = {30, 40, 50, 75, 100, 100, 125, 160, 200, 400}. Meta-research evidence shows that psychology's heterogeneity is inversely correlated with sample size, typically, and simulations confirm that these values of heterogeneity produce the level of correlated heterogeneity observed across dozens of psychology meta-analyses (Stanley, Doucouliagos and Ioannidis, 2022b). To generate random normal heterogeneity, we first convert each estimated correlation to Cohen's *d*, add a random normal deviation with mean zero and standard deviations {.45, .45, .3, .3, .3, .3, .3, .3, .075, .075} and transform these Cohen's *ds* back to correlations.[8]

In the PSB condition, we follow previous studies by assuming that exactly half of the results contained in a meta-analysis have been selected to be statistically significant, while the first random result produced by the other 50% is reported, as it is, statistically significant or not, and included in the meta-analysis (Bartoš, Maier, Wagenmaker, et al., 2023; Bom and Rachinger, 2019; Stanley and Doucouliagos, 2014, 2015; Stanley, Doucouliagos and Ioannidis, 2017). We do not mean to imply that all areas of psychology have such strong selection for statistical significance. Thus, we also report cases of no selection for statistical significance. Table 2 reports the average statistical results across cases where there is 50% publication selection bias and where there is no selection for statistical significance. The average across no publication selection bias (**Het**) and 50% PSB (**PSB**) is likely to better reflect typical areas of psychology, this average is reported in the last row of Table 2, labelled "**PSB & Het Ave.**"

The full details of how we generate 500,000 correlation studies from individual subject data, collectively containing 64 million subjects, are reported in the Supplement and follow previous studies (Stanley, Doucouliagos, Maier et al., 2023; Stanley, Doucouliagos and Havránek, 2023). The central innovations relative to these other simulation studies are: (i) the use of a distribution of sample sizes, rather than a single fixed sample size, (ii) the inclusion of the typical level of heterogeneity, (iii) the infusion of 50% PSB, and (iv) the investigation of the performance of corrections for PSB, PET-PEESE, along with traditional and novel weighted averages.

To generate estimated correlations between two variables, Y_i and X_{1i}, we begin with the simple linear regression:

---

[8] Random heterogeneity added to ρ produce asymmetric, nonnormal, sampling distributions that induce further estimation biases. Conversion to Cohen's *d* avoids this added source of bias. Generating heterogeneity through random variations to X_1's regression coefficient, β_1 of Eq. (9), below, produces approximately the same overall results.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \tag{11}$$

For the sake of transparency and simplicity, we assume that $\beta_0 = \beta_1 = 1$. X_{1i} & ε_i are forced to be independently and normally distributed. We generate ε_i as N(0,1), and X_{1i} as N(0, .25), where .25 is the variance for ρ = 0.447. After which, eq. (11) determines the values of Y_i. Next, the bivariate regression, eq. (11), is estimated along with the *t*-value of the estimated regression coefficient β_1. This *t*-value is converted to a correlation by eq. (4). With these simple data generating processes, we know that the population variance of Y_i is 1.25, because it is the sum of X_{1i} and ε_i's variances. The correlation squared must then be equal to the ratio of the variance of $\beta_1 X_{1i}$ to the variance of Y_i. That is, ρ^2 = .25/1.25 =1/5, making $\rho = \sqrt{\frac{1}{5}}$ or 0.447. 0.447 is the approximate size of the 90th percentile of correlations in psychology as estimated by random effects. In other simulation experiments, we set ρ approximately equal to the median effect size (ρ = sqrt(1/17) = .243) by generating X_{1i} as a random normal, N(0,1), divided by 4 and a 'small' effect size (ρ =sqrt(1/82)=.110) by dividing random N(0,1) by 9. Doing so makes X_{1i}'s variance equal to 1/16 and 1/81, respectively, while leaving the error variance at one.

For each correlational study, all data are generated from eq. (11), this regression is estimated, then r is calculated from eq. (4), S_1^2 is calculated from eq. (2), and S_2^2 from eq. (3). Each of these steps are repeated 50 times to represent *one* meta-analysis.[9] From these 50 randomly generated estimated correlations, RE and the UWLS weighted averages are calculated. For each of 10,000 randomly generated meta-analysis, RE's and UWLS' biases, square roots of the mean squared errors (RMSE), and confidence intervals are calculated and then averaged across these 10,000 replications. See the Supplement for the simulation code. Table 1 reports the results of these simulations using both versions of *r*'s variance—eq. (2) and eq. (3). S_2^2 consistently produces twice the bias as S_1^2. Table 1 also shows that S_2^2 generates larger mean squared errors and inferior coverage (*i.e.*, coverage rates that are often much different than their nominal 95% level). Several lessons can be drawn from this simulation study.

[9] In psychology, the average number of estimated correlations per meta-analysis is 55 (Stanley *et al*., 2018). The biases of correlations are largely independent of the number of correlations (*k*) meta-analyzed. In contrast, the sample size (*n*) of the primary study used to calculate correlations is a very important determinant of this bias, as meta-analysis of correlations suffers from small-sample bias (Stanley, Doucouliagos, Maier et al., 2023).In these simulations, we assume that the distribution of sample sizes reflects what is typically seen in psychology.

5. Results and Discussion

First, Table 1 confirms that the conventional meta-analysis formula of correlations' variance, S_2^2 (Borenstein *et al.,* 2009), should not be used in meta-analysis when an obvious and simple alternative is always available, S_1^2. In all cases, the biases, MSE and CIs are better when S_1^2 is used rather than S_2^2. Thus, by simply not squaring the numerator of correlation's variance formula, the biases and MSEs of conventional meta-analysis can be notably reduced, and the CIs improved. However, when there is no publication bias, the biases of conventional random effects are little more than rounding error ($\leq$ .01). When there is notable selection for statistical significance (i.e., publication selection bias), biases of all simple meta-analysis methods can be substantial. This is especially problematic for more than half of psychological research where effect sizes are small (ρ =.11; .243). With notable publication bias, conventional random-effects meta-analyses of 50 correlations are virtually certain to be falsely positive (that is, to be statistically significant when the correlation is, in fact, zero)—see Table 2.[10] The publication bias of RE is especially pernicious, when research synthesis is needed most: small correlations. For these, the bias of RE is likely to be as large as the true population correlation or nearly so, and RE is likely to falsely suggest a genuine effect where there is none (see Table 2).

Table 2 reports two further meta-analysis estimators, $\mathbf{UWLS_{+3}}$ and **REz,** shown by (Stanley, Doucouliagos and Havránek, 2023) and (Stanley, Doucouliagos, Maier et al., 2023) to outperform conventional meta-analyses of correlations and partial correlations and to reduce their small-sample biases to negligibility. Table 2's simulations confirm that this remains the case even when meta-analyses have a typical distribution of sample sizes and heterogeneity, see the top two thirds of Table 2 and compare them to Table 1. However, as expected, both $\mathbf{UWLS_{+3}}$ and **REz** (along with RE and UWLS) have notable biases when there is 50% PSB.

Note that $\mathbf{UWLS_{+3}}$ generally has better statistical properties than **REz**. This is especially clear when the two most realistic cases, **Het & PSB**, are averaged—see the last row, labelled "**PSB & Het Ave**," in Table 2. Like conventional meta-analyses, $\mathbf{UWLS_{+3}}$ and **REz** have unacceptable

[10] Table 2 reports that the type I error of the random effects estimates that are based on the *z*-transformation, **REz**, to be 99.85%. In all cases, **REz** has better statistical properties than the conventional RE of correlations. Compare $\mathbf{RE_2}$ in Table 1 to **REz** in Table 2.

biases and type I errors when there is notable publication bias. Thus, if publication selection bias is a genuine risk, some PSB accommodation or correction should be used.

In Section 3.3, above, we discussed PET-PEESE as a frequently employed method to reduce publication bias in psychology. Researchers have questioned the validity of PET-PEESE and related meta-regression corrections for publication bias (based on the Egger regression) because SE can be correlated with effect size in the absence of publication bias (Egger et al. 1997; Moreno et al., 2009; Pustejovsky and Rodgers, 2019). Thus, a surprising finding is that, even for correlations, where the correlation with SE in the absence of publication bias is mechanical, PET-PEESE works well to reduce PSB and type I errors when there is publication bias – see the columns for **PET-PEESE** and **PPz** at the bottom of Table 2. Average bias of **PET-PEESE** is only about .01, approximately 4 times smaller than conventional random-effects' biases (using either correlations or Fisher's *z* transformation).

Despite the correlation between SE and *r* in the PET-PEESE meta-regressions, eqs. (7) and (8), PET-PEESE has very good statistical properties. Especially relevant, note that PET's type I errors are always within their nominal levels, whether or not there is publication bias. However, PET-PEESE is not perfect and can be improved through the Fisher's *z* transformation because *z* is not correlated to its SE. **PPz** reports the statistical properties of first converting correlations to *z*, calculating PET-PEESE in terms of Fisher's *z*, and lastly converting PET-PEESE in terms of *z* back to a correlation. On average, **PPz** has smaller bias, MSE, type I errors, and better CIs than **PET-PEESE** of correlations. However, there is a potential problem with using **PPz** in the place of **PET-PEESE**. **PPz** is downwardly biased for small correlations (.11), and this is a rather crucial effect size range as Cohen's guidelines suggest that anything less than .1 is 'trivial' or scientifically 'null' (Cohen, 1990). In contrast, **PET-PEESE** is never downwardly bias, and its upward biases are trivial (< .01) for small 'true' effect sizes. When analyzing effects that may be null or trivial, it would, therefore, be better to use **PET-PEESE** of correlations but to rely on **PPz** in other cases. For the sake of robustness, we recommend that researchers report both.

Surprisingly, across the two most representative research conditions, **Het & PSB** (that is, heterogeneity without PBS and heterogeneity with 50% PSB, respectively), **UWLS$_{+3}$** has the smallest average RMSE. Yet, **UWLS$_{+3}$** does not correct for PSB, explicitly. This relatively small RMSE is not a justification for only using **UWLS$_{+3}$** when PSB is suspected. While its RMSE is as good or better than the alternatives, it also has unacceptable Type I errors (.9957) with 50% PSB. Because PSB can rarely be ruled out either *a priori* or though tests of PSB (as they all tend to have low power), **PET-PEESE** should be routinely reported along with **UWLS+3** and the more conservative results emphasized.

We also extend these findings about simple bivariate correlations to partial correlations, following Fisher's (1921) observation that what works for correlations works for partial correlations. Stanley, Doucouliagos and Havránek (2023) demonstrates how the meta-analysis of partial correlations suffer from the same small-sample biases as do bivariate correlations and that the methods that accommodate these small-sample biases (i.e., **REz** and **UWLS+3**) are equally effective in the meta-analysis of partial correlations. Here, we extend the same simulation design that produced Table 2 to partial correlations.

To do so, we first generate the original correlational data by adding an independent variable, X_{2i}, to eq. (9).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{12}$$

We make the same assumptions about the distribution of these variables and further assume that $\beta_2 = 1$ and force X_{2i} to be independently and normally distributed, N(0,1). Otherwise, this simulation design is identical with the simulations reported above and uses the same GAUSS code after replacing eq. (9) with eq. (10) (see the Supplement). The results for partial correlations (see Table S2 in the Supplement) are virtually the same as those reported in Table 2 and those found in Stanley, Doucouliagos and Havránek (2023). Thus, as Fisher correctly noted more than a century ago, what is true for correlations is also true for partial correlations once degrees of freedom are properly adjusted.

6. Conclusions

Conventional meta-analyses of correlations are biased, even under ideal conditions. However, to isolate and to document these small-sample biases, Stanley, Doucouliagos, Maier et al. (2023) assumed that all studies in a meta-analysis had the same sample size and that there is no selection for statistical significance (i.e., no PSB). The purpose of this paper is to investigate the statistical properties of conventional meta-analysis methods under typical conditions widely seen in correlational research in psychology. We find that these small-sample biases remain although they are, for the most part, negligibly small except when some results are selected for their statistical significance.

When some results are selected for their statistical significance, PET-PEESE has little if any notable bias (< .02), and tests of the correlation's statistical significance (PET) maintain their nominal type I errors. This is an especially surprising finding as estimated correlations are mechanically correlated with their standard errors, though inversely so, in the absence of PSB. This correlation is seen by some to be a disqualifying condition for the application of PET-PEESE and the related Egger regression to meta-analysis (Moreno et al., 2009; Pustejovsky and Rodgers, 2019). PET-PEESE is a notable improvement over random-effects (RE) even when averaged across research areas where there is PSB and where there is no PSB (see the last row of Table 2). With PSB, RE can have biases as large as the population mean correlation it is estimating, and RE is virtually certain (99.85%) to falsely identify statistically significant correlations that do not exist (see Table 2, Type I errors).

This study corroborates other findings of Stanley, Doucouliagos, Maier et al. (2023). They show that the conventional formula for the variance of correlations, $S_2^2 = {(1-r^2)^2}/{(n-1)}$ (Borenstein et al., 2009), should never be used in meta-analysis as it is statistically dominated in all cases by a simpler formula, $S_1^2 = {(1-r^2)}/{(n-2)}$, that does not square the dominator.

We also show that a new simple weighted average, **UWLS+3**, statistically dominates RE whether or not correlations are first transformed to Fisher's *z* (Table 2). This simple correction for small-sample bias, **UWLS+3**, adjusts the degrees of freedom and emerges as the preferred meta-analysis estimator in the absence of PSB. With PSB, PET-PEESE, using either correlations or Fisher's *z*, has the best statistical properties under typical correlational research conditions. Unless publication selection bias can be ruled out *a priori* (for example, in a meta-analysis of pre-registered replications), we recommend researchers report PET-PEESE and emphasize whichever meta-analysis results (from **PET-PEESE**, **REz** or **UWLS+3**) are the more conservative.

Needless to say, there are limitations to our findings. They apply fully only to the specifications that we simulate, which assume that meta-analyses have the typical conditions seen widely across correlational studies in psychology, more or less. However, not all meta-analyses involve 'typical' correlational research. In particular, if all studies use small samples ($n \leq 50$ or $n \leq 100$), small-sample biases will generally be larger and PET-PEESE is no longer valid as there will be too little variation in SE, and, as a result, PET-PEESE will produce unreliable estimates. When there is little variation in *SE*, the independent variable in the PET-PEESE meta-regression will have little useful information with which to estimate their slope coefficient. If there is insufficient information to estimate the slopes accurately, the estimates of the intercepts will be equally unreliable. In meta-analyses with little variation in SE, one should not employ PET-PEESE (Stanley, 2017).[11]

To prevent any undue influence from one or a few overly influential effects, meta-analysts should always use influence statistics (also called leverage points or, incorrectly, 'outliers') to identify and remove such overly influential studies regardless of their cause. The criterion and method used to identify leverage points can be stated in a pre-analysis plan. Without the identification and removal of highly influential effect sizes, any meta-analysis result can be highly skewed towards simple coding/transcription/transformation errors or, in rare cases, fraud.[12]

In summary, a simple adjustment to degrees of freedom can reduce the small-sample bias of the meta-analysis of correlations, and publication selection biases are almost fully corrected by PET-PEESE under typical conditions seen widely across correlational studies in psychology.

---

[11] However, it needs to be emphasized that meta-analyses in psychology typically have sufficient variation in sample size and SE to allow the PET-PEESE models to be reliably estimated. Across 600 psychology meta-analyses, the typical (median) ratio between the smallest sample size and the largest is a factor of 20 (Bartoš, Maier & Wagenmakers et al., 2022). Thus, the typical distribution of sample sizes across psychology is wider than the distribution of sample sizes used in this paper's simulations. In those rare cases where there is little variation among samples sizes (e.g., $n \leq 100$, for all studies), we recommend Bayesian model averaging that lets the research record, itself, decide on the appropriate weights (Bartoš, Maier & Wagenmakers et al., 2023).

[12] An illustrative example comes from Kivikangas' *et al.* (2021) meta-analysis of the correlation between moral foundations and political orientation. In this area of research, one study stands out, Graham *et al.* (2011). It uses data from the YourMorals.org, which required the volition of over 200,000 individual subjects. Including this one study doubles the mean effect size, as Graham *et al*. (2011) reports both the largest correlations and the largest sample by more than an order of magnitude. This study's large effect size is probably not an error and clearly not fraud. As Kivikangas *et al.* (2021) argue the large effect was likely the result of self-selection to participate by those with the more extreme political orientations. Regardless of cause, such overly influential studies need to be omitted or accommodated through moderator analysis, just as Kivikangas *et al*. (2021) did.

Highlights

What is already known?

- All inverse-variance weighted meta-analyses of correlations are biased.
- Dozens, perhaps hundreds, of meta-analyses of correlations are conducted each year.

What is new?

- We investigate the statistical properties of alternative meta-analysis estimators of the population correlation coefficient with simulations that closely match typical research conditions widely seen across correlational studies in psychology.
- We explore a novel correction, $UWLS_{+3,}$ that reduces these small-sample biases to scientific negligibility.
- $UWLS_{+3}$ is the unrestricted weighted least squares weighted average that adjusts degrees of freedom to effectively eliminate small-sample bias. It is less biased than random effects calculated on correlations or Fisher's *z* whether there is publication selection bias or not.
- Despite the mechanical correlation between estimated correlations and their standard errors, PET-PEESE effectively removes publication selection bias under the typical research conditions widely found across correlational studies in psychology.

References:

Bartoš F, Maier M, Shanks DR, Stanley TD, Sladekova M, Wagenmakers EJ (2023). Meta-analyses in psychology often overestimate evidence for and size of effects. *Royal Society: Open Science*. 05 July, https://doi.org/10.1098/rsos.230224

Bartoš F, Maier M, Wagenmakers EJ, Doucouliagos H, Stanley TD. (2023). Robust Bayesian meta-analysis: Model averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*. 14:99–116.

Bartoš F, Maier M, Wagenmakers EJ, Nippold F, Doucouliagos H, Ioannidis JPA, Otte WM, Sladekova M, Fanelli D, Stanley TD. (2022). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. Preprint at https://doi.org/10.48550/arXiv.2208.12334 .

Bom PRD, Rachinger H. (2019) A kinked meta-regression model for publication bias correction. *Research Synthesis Methods* 10:497e514.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons.

Carter EC, Kofler LE, Forster DF, McCullough ME. (2015) A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General*, 144, 796–715.

Carter EC, Schonbrodt FD, Gervais WM, Hilgard J. (2019). Correcting for bias in psychology: a comparison of meta-analytic methods. *Advances in Methods and Practices of Psychological Science* 2(2):115e44.

Cohen J. (1990). Things I learned (so far). *Am Psychol* 4 5:1304–1312.

Doucouliagos, C(H). and Stanley, T.D. (2013) Theory competition and selectivity: Are all economic facts greatly exaggerated? *Journal of Economic Surveys*, 27: 316-39.

Eastwick PW, Luchies LB, Finkel EJ, Hunt ll. (2014). The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin*, 140: 623–665.

Egger M, Smith GD, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal,* 315:629–634.

Fisher RA. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.

Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, Ditto PH. (2011). Mapping the moral domain. *Journal of Personality & Social Psychology*, *101*(2), 366–385.

Gustafson RL. (1961). Partial correlations in regression computations. *Journal of the American Statistical Association*, 56(294): 363–367.

Hagger MS, Chatzisarantis NLD, Alberts H, et al. (2016). A multi-lab preregistered replication of the ego-depletion effect. *Perspect Psychol Sci.* 11:546–573.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Irsova Z, Bom PRD, Havranek T, Rachinger H. (2023). Spurious precision in meta-analysis. MetaArXiv. February 23. doi:10.31222/osf.io/3qp2w.

Klein RA, Vianello M, Hasselman F, Alper S, Aveyard M, Axt JR, Nosek BA. (2018). *Many Labs 2: Investigating variation in replicability across sample and setting*, 1(4): 443–490.

Kivikangas JM, Fernández-Castilla B, Järvelä S, Ravaja N, Lönnqvist J-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin*, *147*(1), 55–94. https://doi.org/10.1037/bul0000308

Kvarven A, Strømland E, Johannesson M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nat Hum Behav*; 4:659–663.

Moreno SG, Sutton AJ, Ades AE, et al. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol*. 9:2.

Pustejovsky JE, Rodgers MA. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*. 10:57–71.

Stanley TD. (2005). Beyond publication bias. *Journal of Economic Surveys*, 19: 309–345.

Stanley TD. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychology and Personality Science*, 8: 581–591.

Stanley TD, Carter E, Doucouliagos H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*. 144:1325–1346.

Stanley TD and Doucouliagos H. (2014). Meta-regression approximations to reduce publication selection bias, *Research Synthesis Methods* 5,60–78.

Stanley TD, Doucouliagos H. (2015). Neither fixed nor random: weighted least squares meta-analysis. *Statistics in Medicine*, 34:2116e27.

Stanley TD, Doucouliagos H. (2017). Neither fixed nor random: Weighted least squares meta-regression analysis, *Research Synthesis Methods*, 8:19–42.

Stanley TD, Doucouliagos H. (2022). Harnessing the power of excess statistical significance: Weighted and iterative least squares. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000502

Stanley TD, Doucouliagos H. (2023). Correct standard errors can bias meta-analysis. *Research Synthesis Methods*, 14: 515–519.

Stanley TD, Doucouliagos H, Ioannidis JPA. (2017). Finding the power to reduce publication bias, *Statistics in Medicine*, 36: 1580–1598.

Stanley TD, Doucouliagos H, Ioannidis JPA. (2022a). Retrospective median power, false positive meta-analysis and large-scale replication. *Research Synthesis Methods*, 13, 88–108.

Stanley TD, Doucouliagos H, Ioannidis JPA. (2022b). Beyond random effects: When small-study findings are more heterogeneous. *Advances in Methods and Practices of Psychological Science*, 5, July-September, 1–11.

Stanley TD, Doucouliagos H, Havránek T. (2023). Meta-analyses of partial correlations are biased: Detection and solutionhttps://www.econstor.eu/bitstream/10419/270940/1/pcc.pdf

Stanley TD, Doucouliagos H, Maier, M, Bartoš F. (2023). Correcting bias in the meta-analysis of correlations. psyarxiv.com: DOI:10.31234/osf.io/cjz56.

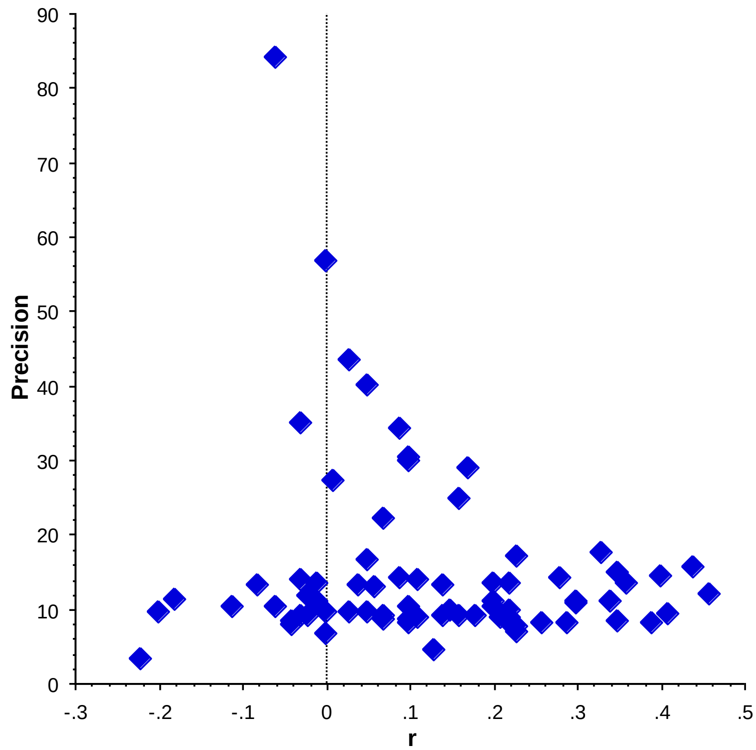

FIGURE 1: A plot of the earnings-romance correlations for women against their precision, $1/S_1$, on the vertical axis (Eastwick, Luchies & Finkel *et al*., 2013).

Table 1: The meta-analyses of correlations (RE and UWLS) using different formulas for the correlation variance

Design			Bias				Coverage				RMSE			
Het/PSB	ρ	I^2	RE_2	RE_1	$UWLS_2$	$UWLS_1$	RE_2	RE_1	$UWLS_2$	$UWLS_1$	RE_2	RE_1	$UWLS_2$	$UWLS_1$
None	.447	.1058	.0097	.0042	.0103	.0042	.8595	.9572	.8362	.8925	.0142	.0110	.0147	.0110
None	.243	.1069	.0065	.0028	.0069	.0030	.9261	.9562	.9177	.9318	.0139	.0123	.0140	.0123
None	.110	.1079	.0028	.0010	.0030	.0011	.9530	.9603	.9483	.9493	.0132	.0126	.0131	.0125
Average			.0063	.0027	.0067	.0028	.9129	.9579	.9007	.9245	.0138	.0120	.0139	.0120
Het	.447	.6023	.0075	.0007	.0197	.0057	.8991	.9357	.7676	.8960	.0192	.0173	.0254	.0162
Het	.243	.6491	.0036	-.0009	.0138	.0039	.9247	.9331	.8928	.9377	.0228	.0216	.0244	.0192
Het	.110	.6638	.0016	-.0006	.0068	.0019	.9276	.9308	.9404	.9511	.0240	.0229	.0225	.0201
Average			.0042	-.0003	.0135	.0038	.9171	.9332	.8669	.9283	.0220	.0206	.0241	.0185
PSB	.447	.5668	.0190	.0113	.0260	.0124	.7573	.8686	.5988	.8019	.0250	.0194	.0301	.0190
PSB	.243	.6171	.0527	.0451	.0478	.0363	.2305	.3304	.2846	.4439	.0562	.0488	.0510	.0400
PSB	.110	.6879	.0955	.0894	.0819	.0725	.0187	.0226	.0368	.0560	.0982	.0919	.0842	.0748
Average			.0557	.0486	.0519	.0404	.3355	.4072	.3067	.4339	.0598	.0534	.0551	.0446
PB & Het Ave			.0300	.0242	.0327	.0221	.6263	.6702	.5868	.6811	.0409	.0370	.0396	.0316

Notes: **HET/PSB** describes different assumed conditions. With **PSB,** the simulations force both heterogeneity and 50% of the study results to be selected for statistical significance i.e. publication bias, **Het** assumes only heterogeneity, and **None** allows neither. ρ is the 'true' population correlation. **Bias** is the difference between the meta-analysis estimate calculated from 50 estimated correlation coefficients and averaged across 10,000 replications. **RMSE** is the square root of the mean squared error. **Coverage** is the proportion of 10,000 meta-analyses' 95% confidence intervals that contain ρ. **RE** is the random-effect's estimate of the mean, and **UWLS** is the unrestricted weighted least squares' estimate of the mean. The subscripts (1 and 2) refer to the use of either correlation variance, S_1^2, from eq. (2) or S_2^2 from eq. (3) to calculate RE's and UWLS' weighted averages. $\mathbf{I^2}$ is a widely known, relative measure of heterogeneity.

Table 2: **RE$_z$**, **UWLS$_{+3}$**, and **PET-PEESE** meta-analyses of correlations

Design			**Bias**				**Coverage**				**RMSE**			
Het/PSB	ρ	**I^2**	**UWLS+3**	**REz**	**PP**	**PPz**	**UWLS+3**	**REz**	**PP**	**PPz**	**UWLS+3**	**REz**	**PP**	**PPz**
None	.447	.1042	.0001	.0016	.0128	.0001	.9480	.9555	.8674	.9469	.0101	.0103	.0198	.0153
None	.243	.1048	.0000	.0010	.0083	.0001	.9469	.9554	.9266	.9472	.0119	.0121	.0199	.0180
None	.110	.1076	.0002	.0006	.0033	-.0010	.9512	.9585	.9435	.9439	.0124	.0126	.0217	.0221
Average			**.0001**	**.0011**	**.0081**	**-.0003**	**.9487**	**.9565**	**.9125**	**.9460**	**.0115**	**.0116**	**.0205**	**.0185**
Type I error rate							**.0250**	**.0206**	**.0248**	**.0265**				
Het	.447	.6019	.0012	-.0011	.0274	-.0008	.9561	.9378	.8170	.9719	.0153	.0172	.0343	.0210
Het	.243	.6500	.0013	-.0003	.0192	-.0008	.9562	.9376	.9347	.9736	.0187	.0212	.0321	.0254
Het	.110	.6635	.0003	-.0007	.0056	-.0066	.9543	.9334	.9684	.9777	.0201	.0228	.0350	.0369
Average			**.0009**	**-.0007**	**.0174**	**-.0028**	**.9555**	**.9363**	**.9067**	**.9744**	**.0180**	**.0204**	**.0338**	**.0278**
Type I error rate							**.0198**	**.0328**	**.0074**	**.0094**				
PSB	.447	.5650	.0080	.0093	.0198	-.0073	.9166	.8954	.8912	.9598	.0166	.0184	.0284	.0215
PSB	.243	.6160	.0341	.0445	.0141	-.0060	.5301	.3427	.9463	.9641	.0380	.0483	.0292	.0256
PSB	.110	.6877	.0715	.0888	.0057	-.0208	.0699	.0239	.8260	.8648	.0739	.0914	.0611	.0639
Average			**.0379**	**.0476**	**.0132**	**-.0114**	**.5055**	**.4207**	**.8878**	**.9296**	**.0429**	**.0527**	**.0396**	**.0370**
Type I error rate							**.9957**	**.9985**	**.0169**	**.0104**			**Type**	**I error rat**
PSB & Het Ave			**.0194**	**.0234**	**.0153**	**-.0071**	**.7305**	**.6785**	**.8973**	**.9520**	**.0304**	**.0365**	**.0367**	**.0324**

Notes: **HET/PSB** describes different assumed conditions. With **PSB,** the simulations force both heterogeneity and 50% of the study results to be selected for statistical significance i.e. publication bias, **Het** assumes only heterogeneity, and **None** allows neither. ρ is the 'true' population correlation. **Bias** is the difference between the meta-analysis estimate calculated from 50 estimated correlation coefficients and averaged across 10,000 replications. **RMSE** is the square root of the mean squared error. **Coverage** is the proportion of 10,000 meta-analyses' 95% confidence intervals that contain ρ. Type I errors, by definition, must assume that ρ =0, and thereby only be reported once for each design condition. **RE** is the random-effect's estimate of the mean, and **UWLS** is the unrestricted weighted least squares' estimate of the mean. **UWLS+$_3$**, as discussed in text, is the unrestricted weighted least squares meta-average with 3 additional degrees of freedom, **REz** is the random effects estimate of the mean correlation after being transformed back from Fisher's z, **PP** is PET-PEESE, and **PPz** is the PET-PEESE that uses the Fisher's z transformation. **I^2** is a widely known, relative measure of heterogeneity.

IES Working Paper Series

2023

1. Josef Bajzik, Tomáš Havránek, Zuzana Iršová, Jiří Novák: *Are Estimates of the Impact of Shareholder Activism Published Selectively?*
2. Klára Kantová: *Ex-Prisoners and the Labour Market in the Czech Republic*
3. Theodor Petřík, Martin Plajner: *Concurrent Business and Distribution Strategy Planning Using Bayesian Networks*
4. Tijmen Tuinsma, Kristof De Witte, Petr Janský, Miroslav Palanský, Vitezslav Titld: *Effects of Corporate Transparency on Tax Avoidance: Evidence from Country-by-Country Reporting*
5. Zuzana Irsova, Pedro R. D. Bom, Tomas Havranek, Heiko Rachinger: *Spurious Precision in Meta-Analysis*
6. Vojtěch Mišák: *Does Heat Cause Homicides? A Meta-Analysis*
7. Fan Yang: *The Impact of Regulatory Change on Hedge Fund Performance*
8. Boris Fisera: *Distributional Effects of Exchange Rate Depreciations: Beggar-Thy-Neighbour or Beggar-Thyself?*
9. Salim Turdaliev: *Powering Up Cleaner Choices: A Study on the Heterogenous Effects of Social Norm-Based Electricity Pricing on Dirty Fuel Purchases*
10. Kseniya Bortnikova: *Beauty and Productivity in Academic Publishing*
11. Vladimír Benáček, Pavol Frič: *Ossified Democracy as an Economic Problem and Policies for Reclaiming its Performance*
12. Petr Janský, Miroslav Palanský, Jiří Skuhrovec: *Public Procurement and Tax Havens*
13. Katarzyna Bilicka, Evgeniya Dubinina, Petr Janský: *Fiscal Consequences of Corporate Tax Avoidance*
14. Evžen Kočenda, Shivendra Rai: *Drivers of Private Equity Activity across Europe: An East-West Comparison*
15. Adam Geršl, Barbara Livorová: *Does Monetary Policy Reinforce the Effects of Macroprudential Policy*
16. Tersoo David Iorngurum: *Method versus cross-country heterogeneity in the exchange rate pass-through*
17. T. D. Stanley, Hristos Doucouliagos, Tomas Havranek: *Meta-Analyses of Partial Correlations Are Biased: Detection and Solutions*
18. Samuel Fiifi Eshun, Evžen Kočenda: *Determinants of Financial Inclusion in Africa and OECD Countries*
19. Matej Opatrny, Tomas Havranek, Zuzana Irsova, Milan Scasny: *Publication Bias and Model Uncertainty in Measuring the Effect of Class Size on Achievement*
20. Soňa Sivá: *Effects of Government Interventions on Bank Performance*
21. Oleg Alekseev, Karel Janda, Mathieu Petit, David Zilberman: *Impact of Raw Material Price Volatility on Returns in Electric Vehicles Supply Chain*

22. Karel Janda, Barbora Schererova, Jan Sila, David Zilberman: *Graph Theory Approach to Prices Transmission in the Network of Commonly Used Liquid Fuels*
23. Yermone Sargsyan, Salim Turdaliev, Silvester van Koten: *The Heterogeneous Effects of Social Cues on Day Time and Night Time Electricity Usage, and Appliance Purchase: Evidence from a Field Experiment in Armenia*
24. Jan Sila, Evzen Kocenda, Ladislav Kristoufek, Jiri Kukacka: *Good vs. Bad Volatility in Major Cryptocurrencies: The Dichotomy and Drivers of Connectedness*
25. Zuzana Irsova, Hristos Doucouliagos, Tomas Havranek, T. D. Stanley: *Meta-Analysis of Social Science Research: A Practitioner's Guide*
26. Diana Kmetkova, Milan Scasny, Iva Zverinova, Vojtech Maca: *Exploring the Link Between Diet and Sustainability in Europe: A Focus on Meat and Fish Consumption*
27. Fisnik Bajrami: *The Impact of Dollarisation on Economic Growth, Investment, and Trade*
28. Miroslav Svoboda, Michael Fanta, Jan Mošovský: *Effectiveness of Car Scrappage Schemes: Comparative Analysis of European Countries*
29. Nicolas Fanta, Roman Horvath: *Artificial Intelligence and Central Bank Communication: The Case of the ECB„*
30. Karel Janda, Jan Sila, David Zilberman: *Fueling Financial Stability: The Financial Impact of U.S. Renewable Fuel Standard*
31. Anna Pavlovova: *High-Frequency Groceries Prices: Evidence from Czechia*
32. Kumar Chandrakamal Pramod Kumar: *Less-cash or more-cash? Determinants and trends of currency in circulation in a panel of 17 economies*
33. Javier Garcia-Bernardo, Petr Janský:*Profit Shifting of Multinational Corporations Worldwide*
34. T. D. Stanley, Hristos Doucouliagos, Tomas Havranek: *Reducing the Biases of the Conventional Meta-Analysis of Correlations*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz