Original Article

Research Synthesis Methods

Received 12 July 2012, Revised 16 July 2013, Accepted 21 July 2013 Published online 3 September 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1095

Meta-regression approximations to reduce publication selection bias

T. D. Stanley[a*†] and Hristos Doucouliagos[b]

Publication selection bias is a serious challenge to the integrity of all empirical sciences. We derive meta-regression approximations to reduce this bias. Our approach employs Taylor polynomial approximations to the conditional mean of a truncated distribution. A quadratic approximation without a linear term, precision-effect estimate with standard error (PEESE), is shown to have the smallest bias and mean squared error in most cases and to outperform conventional meta-analysis estimators, often by a great deal. Monte Carlo simulations also demonstrate how a new hybrid estimator that conditionally combines PEESE and the Egger regression intercept can provide a practical solution to publication selection bias. PEESE is easily expanded to accommodate systematic heterogeneity along with complex and differential publication selection bias that is related to moderator variables. By providing an intuitive reason for these approximations, we can also explain why the Egger regression works so well and when it does not. These meta-regression methods are applied to several policy-relevant areas of research including antidepressant effectiveness, the value of a statistical life, the minimum wage, and nicotine replacement therapy. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: meta-regression; publication selection bias; systematic reviews, truncation

1. Introduction

> Many other commentators have addressed the issue of publication bias. . . . All agree that it is a serious problem—Begg and Berlin (1988, p. 421).

The bias that arises from the preferential reporting of statistically significant or 'positive' scientific results has long been a focus and concern of statisticians (Sterling, 1959; Rosenthal, 1979; Hedges and Olkin, 1985; Begg and Berlin, 1988; Sterling, Rosenbaum and Weinkam, 1995; Copas, 1999; Senn, 2008; Mandel and Rinott, 2009; and Rücker *et al*., 2011, to mention a few). This 'publication bias' is widely recognized to exaggerate the effectiveness of pharmaceuticals (Friedman, 2003; Cary, 2008; Turner *et al*., 2008). Publication bias is a misnomer; 'reporting bias' would be a more accurate reflection of this threat to scientific validity. Because the preference for statistical significance is widely known among researchers, they will tend to select statistically significant findings even in their unpublished working papers and theses. Others have found publication selection to be widespread in the natural sciences and economics (Sterling and Weinkam, 1995; Doucouliagos and Stanley, 2013).

As shown in the succeeding text, the reported values of a statistical life are highly skewed and exaggerated (Bellavance *et al*., 2009), and nearly, the entire left side of the results from clinical trials of antidepressants is missing from the published record (Turner *et al*., 2008). How can health care providers or policy makers sensibly correct for publication selection? We seek a practical solution to this widespread problem in social science and medical research.

To minimize publication selection bias, the leading medical journals require the prior registration of clinical trials as a condition of their later publication (Krakovsky, 2004). Nonetheless, a recent systematic review found that publication selection is quite common in medical research (Hopewell *et al*., 2009). Without some way to correct or minimize this bias, the validity of science itself comes into question (Lehrer, 2010).

[a]*Economics, Hendrix College, 1600 Washington St., Conway, AR 72032, USA*
[b]*School of Accounting, Economics, and Finance and Alfred Deakin Research Institute, Deakin University, 221 Burwood Highway, Burwood, 3125, Victoria, Australia*
**Correspondence to: Tom D. Stanley, Professor of Economics, Hendrix College, 1600 Washington St., Conway, AR 72032, USA.*
[†]Email: Stanley@hendrix.edu

In this paper, we derive a practical solution to the exaggerated scientific record. Simple meta-regression analysis (MRA) can greatly reduce publication selection bias. Following the seminal work of Begg and Berlin (1988) and Copas (1999), we recognize that it may not be feasible to estimate all the needed parameters of a fully specified statistical model of publication selection. 'It is difficult to conceive of a correction methodology which would be universally credible' (Begg and Berlin, 1988, p. 440). Nonetheless, we identify an approximate meta-regression model from considerations of limiting cases and a quadratic Taylor polynomial for the expected value of a truncated normal distribution. Furthermore, this meta-regression model easily accommodates research heterogeneity from different methods, data, populations, controls, and so on and can thereby distinguish publication selectivity from more substantive research differences.

The purpose of this paper is to derive statistically and intuitively meta-regression approximations for publication bias, improve them by combining the quadratic meta-regression approximation with the better-known Egger regression intercept (Egger *et al.*, 1997; Stanley, 2008), and offer a new multiple meta-regression model that accommodates differential publication selection that may be related to any number of research dimensions along with systematic heterogeneity and publication bias. This multiple MRA can explicitly distinguish publication bias from funnel asymmetry that may be coincidentally correlated with moderator variables as well as a more complex form of publication selection that depends on other research factors. In the process of deriving and discussing these meta-regression approximations, we explain when and why the Egger regression works well and when it does not (Egger *et al.*, 1997).

Simulations reported here show that a quadratic meta-regression approximation can greatly decrease publication selection bias found in the conventional meta-analytic summary statistics of reported research results. This approach has already been successfully applied to correct highly exaggerated research on the efficiency wage hypothesis (Stanley and Doucouliagos, 2007), antidepressant effectiveness (Moreno *et al.*, 2009b), trade effects of joining the Euro (Havranek, 2010), health care and income (Costa-Font *et al.*, 2011), the relation of foreign investments and taxes (Feld and Heckemeyer, 2011), and the value of a statistical life (VSL) (Doucouliagos *et al.*, 2012).

2. Models of publication selection

2.1. Publication selection as truncation

When all results are selected to be statistically significant in a desirable direction, reported effects may be regarded as 'incidentally' truncated. It is 'incidental' truncation because the magnitude of the reported effect, itself, is not selected but rather some related variable, for example the calculated z-value or t-value (Wooldridge, 2002, p. 552). By referring to the well-known conditional expectation of a truncated normal distribution, it is easy to show that observed effects will depend on the population's 'true' effect, μ, plus a term that reflects selection bias.

$$E(effect_i|truncation) = \mu + \sigma_i \cdot \lambda(c). \quad (1)$$

$\lambda(c)$ is the inverse Mills' ratio, μ is the 'true' effect, which is the expected value of the original distribution, σ_i is the standard error (SE) of the estimated effect, $c = a - \mu/\sigma_i$, and a is the critical value of the standard normal distribution (Johnson and Kotz, 1970, p. 278; Greene, 1990, Theorem 21.2). With selection for directional statistical significance, an estimated effect is more likely to be observed if $effect_i/\sigma_i > a$.

When we replace σ_i in Equation (1) with its sample estimate, SE_i,

$$E(effect_i) = \mu + SE_i \cdot \lambda(c). \quad (2)$$

Equation (2) may be interpreted as a MRA of observed effect on its SE. Unfortunately, $\lambda(c)$ is not generally constant with respect to μ and σ_i, and this complication causes additional difficulty in providing an unbiased corrected estimate.

To clarify the context of this selection problem, we briefly digress. In econometrics, there is the well-known Heckman two-step solution to the analogous problem of sample selection (Heckman, 1979; Wooldridge, 2002; Davidson and MacKinnon, 2004). However, in the empirically tractable case of sample selection, characteristics of the unselected individuals are observed and used to estimate a selection equation, by logit or probit. The estimated values of the inverse Mills' ratio, $\lambda(c)$, from this selection relation are then used to estimate the Heckman regression, which is similar to our MRA in the preceding text. What makes the Heckman approach feasible is the additional information contained in the selection variables that are observed whether the individual is selected or not. We do not have the luxury of extra relevant information in the case of publication selection. In general, nothing is known about the unreported empirical research results. Thus, this well-worn avenue is unavailable for the problem at hand.

Rather than give up altogether, let us approximate the publication bias term, $SE_i \cdot \lambda(c)$, by other means. The inverse Mills' ratio is the normal probability density function, $\phi(c)$, evaluated at $c = a - \mu/\sigma_i$ and divided by one minus its cumulative density, $[1-\Phi(c)]$. As a consequence, this term is a complex function of μ and σ_i. To understand this complexity somewhat better, we take the derivative of Equation (1) with respect to σ_i.

$$\begin{aligned} \partial E(effect_i|truncation)/\partial\sigma_i &= \lambda(c) + \sigma_i \cdot \partial\lambda(c)/\partial\sigma_i \\ &= \lambda(c) + \sigma_i \cdot \partial\lambda(c)/\partial c \cdot (\partial c/\partial\sigma_i). \end{aligned} \quad (3)$$

However, $\partial\lambda(c)/\partial c = \lambda(c)^2 - c\lambda(c)$(Heckman, 1979, p. 159), which gives:

$$\partial E(effect_i|truncation)/\partial\sigma_i = \lambda(c) + (\mu/\sigma_i)\cdot(\lambda(c)^2 - c\lambda(c)). \quad (4)$$

This derivative suggests that the conditional mean is, in general, a rather difficult, nonlinear function of σ_i; thus, some approximation such as the Taylor polynomial (or power series) will need to be employed to estimate the expected empirical relation between a reported estimate and its SE.

$$effect_i = \beta_1 + \sum_{k=1}^{K} \alpha_k SE_i^k + \varepsilon_i. \quad (5)$$

Estimates of β_1 from this Taylor polynomial approximation, Equation (5), will then serve as estimates of the 'true' effect, μ, corrected for publication bias. Econometricians typically employ linear or quadratic approximations in similar applications. In our simulations, in the succeeding text, we investigate quadratic (i.e., K = 2), cubic (i.e., K = 3), as well as linear approximations (i.e., K = 1). However, before we turn to these simulations, we need to make several relevant observations.

There is a rich, 200-year history of constructing limits and approximations for the Mills' ratio, hence, by extension for the inverse Mills' ratio (Laplace, 1812; Johnson and Kotz, 1970). Some of these approximations are in fact power series (Abramowitz and Stegun, 1964). For our application, all of these approximations will involve complex functions of μ/σ_i and thereby involve the very parameter, μ, we wish to estimate. Unfortunately, we find no specific estimation model that can be derived from these approximations. A possible exception is Gordon's (1941) upper bound for the Mills' ratio. When this upper bound is applied to our Equation (1), it gives $E(effect_i|truncation) = a\sigma_i$ as a lower bound. However, our limit cases, especially E(Effect|Selection) in Figure 1, are more informative and useful.

2.2. *Examining limit cases of publication selection*

Examining limit cases reveals how a parabola in SE_i might provide an adequate approximation to the relation between the effect size and its SE. Figure 1 plots 300 randomly generated yet selected effects when there is strict selection of significantly positive effects, and the true effect is one ($\mu = 1$). These randomly generated values come from the same data generating processes used by the simulations reported and discussed in the next section. However, for our present purposes, the limiting cases of publication bias represented by the two lines in Figure 1 are much more informative than any random scatter of selected results. These limit cases give shape to the relationship between the expected reported effect and the SEs. As we discuss in the succeeding text, this shape is known *a priori* from statistical theory.

To understand the shaping forces of these simple lines, first consider the horizontal line, E(Effect|No selection). When all empirical findings are reported with no selection, they will be randomly distributed, by definition, around

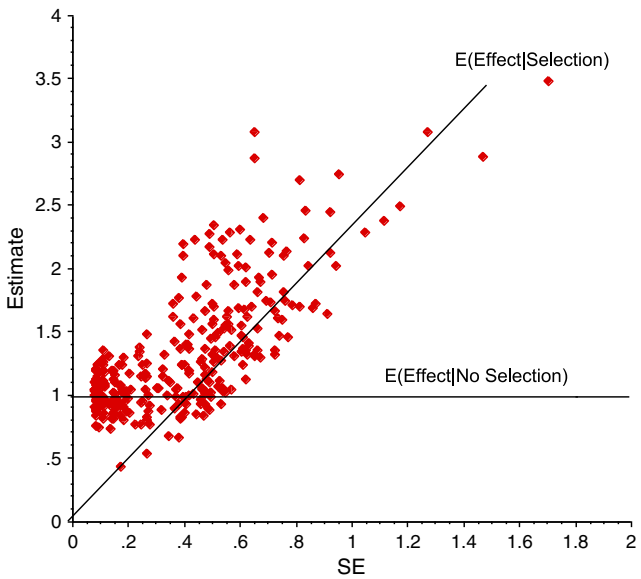

Figure 1. Plots 300 randomly generated yet selected effects (vertical axis) against their standard errors.

the true effect, $\mu = 1$ for this illustration. Without selection, the magnitude of the reported effect will be constant and independent of SE_i, hence, the horizontal line. Next, note the upwardly sloping line in Figure 1, E(Effect|Selection). This second line represents the conditional expectation, Equation (1), when the true effect is zero, $\mu = 0$. This upward-sloping line represents the worst-case scenario for publication bias. The slope of this line will be equal to the inverse Mills' ratio evaluated at the critical value, a. To see this, substitute $\mu = 0$ into Equation (4)—also see Section 2.3. To a greater or lesser extent, these two polar cases shape the reported effects.

Beginning with the most precise studies (those with small SE_i), researchers will find little need to report anything other than the first observed effect. When the true effect is many times larger than the SE, the probability of finding an insignificant effect is virtually zero. Thus, even when there is selection for a statistically positive effect, very precise studies will not be biased, assuming of course that there is some genuine positive effect to begin with. As SE increases, occasionally an estimated effect will not be statistically significant and will need to be re-estimated to become so. Thus, for the 'middle' range of SE, expected observed effects will be gradually pulled up above the horizontal line. Notice the scatter for $0.3 \leq \text{SE} \leq 0.5$ in Figure 1. As SE grows larger still, the standardized true effect, μ/SE_i, will play a weaker and weaker role, whereas the ray from the origin presents a greater attraction for reported effects. In the limit, expected reported effects and their SEs will be linearly related, $\lambda(a)SE_i$.

Thus, a simple thought experiment identifies rather clearly the approximate shape of expected reported effects and their SEs and, thereby, provides plausibility for the meta-regression model explored here. Equally apparent is that the right half of a parabola ($E(effect_i) = \beta_1 + \alpha_2 SE_i^2$) can approximate this relationship. Note further that a parabola will also approximate this relationship when μ is increased or decreased. Changing μ lengthens or shortens the horizontal line segment. Making α_2 smaller in $E(effect_i) = \alpha_2 SE_i^2$ allows for a more gradual increase initially and a wider parabola. Of course, the fit will not be exact, but then, we need only to estimate the minimum of this the parabola (i.e., its vertex).

Our purpose for estimating this relationship between reported effect and its SE is merely to find an adequate corrected estimate of effect, and we know that there will be no publication bias when SE is small, approaching zero. In practice, think of σ_i between 0 and $0.1 \times \mu$. For such small σ_i, every observed effect will be statistically significant; thus, there will be no publication selection bias near 0. In this region, observed effects will not depend on σ_i but rather be randomly distributed around μ. From Equation (1), we know that $f(0) = 0$ in $E(effect_i) = \mu + f(\sigma_i)$. Also, we know from Equation (3) that as σ_i approaches 0 from above $f'(0) \rightarrow 0$ for $\mu \neq 0$. These properties give shape to the relationship between reported effects and their SEs and constrain the form of the approximation that is needed to model this relationship. Figure 1 also makes these points clear. The easiest way to give a Taylor polynomial, these properties is to require that the linear term of a quadratic approximation be omitted from Equation (5); that is, $\alpha_1 = 0$ in Equation (5).

Our simulations, described in the succeeding text, demonstrate that constraining α_1 to be zero in the quadratic approximation of Equation (5) is critical. As discussed previously, very precise estimates will vary around μ and contain negligible publication bias. This observation serves as the starting point for an alternative estimate of the corrected effect—'*Top 10*.' *Top10* is the simple mean of the most precise 10% of the estimates in a research literature and has been shown to reduce publication bias significantly (Stanley *et al*., 2010). Thus, our ideal corrected estimate is where this relation crosses the vertical axis, where SE approaches zero. The trick, of course, is to estimate this intersection well from the statistical results typically reported in empirical studies.

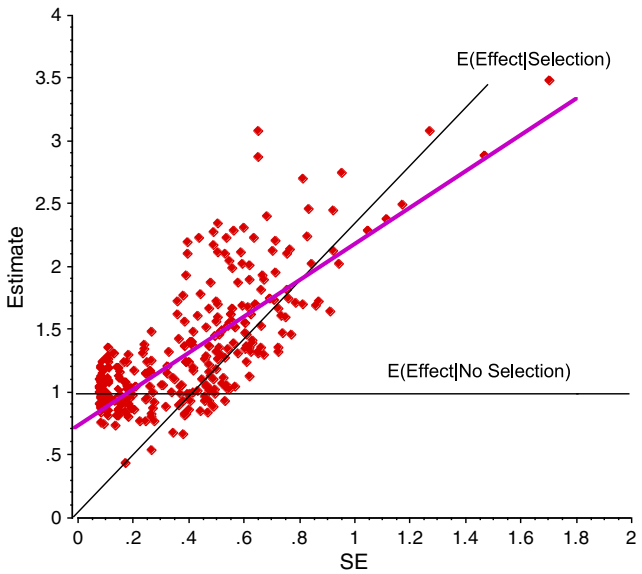

Figure 2. Plots 300 randomly generated yet selected effects (vertical axis) against their standard errors along with the least squares line.

Using a linear approximation to the Taylor polynomial would be one approach, but not a very good one. Previous simulations show that this leads to an underestimate of the true effect when there is an effect (Stanley, 2008), and this is easily seen in Figure 2. Figure 2 places the least squares line (upward sloping with a positive intercept) through this scatter of reported effects, the intercept of which underestimates the true effect by 25%. This illustration is no isolated incident but is robustly confirmed by the simulations reported in the next section. In spite of this bias, there is an important special case where the expected reported effect and its SE will be linearly related.

2.3. *Egger regression and the precision-effect test*

Egger *et al*. (1997) used the linear approximation to this complex relation of reported effect to its SE as a test for the presence of publication bias.

$$effect_i = \gamma_1 + \alpha_1 SE_i + \varepsilon_i. \tag{6}$$

Testing H_0: $\alpha_1 = 0$ in this simple meta-regression model is widely used in medical research to investigate whether a research literature is contaminated by publication selection or more generally 'small-study' bias. This Egger test serves as a valid if low power test for publication selection or small-sample bias (Egger *et al*., 1997; Stanley, 2008). This test is related to the symmetry of the associated funnel graph. A funnel graph is a plot of precision ($1/SE_i$) versus $effect_i$, and it has been widely used in systematic reviews as a visual indicator of publication selection (Egger *et al*., 1997; Duval and Tweedie, 2000; Hopewell *et al*., 2009). Because this MRA contains obvious heteroscedasticity, Equation (6) is almost never estimated using ordinary least squares (OLS), but rather weighted least squares (WLS). WLS can be obtained by dividing the entire Equation (6) by an estimate of the standard deviation of this heteroscedasticity (i.e., SE_i) or by using a WLS routine in any standard statistical package when the weights are specified as $1/SE_i^2$. See the Appendix for the appropriate SPSS and STATA commands.

$$t_i = \alpha_1 + \gamma_1(1/SE_i) + u_i. \tag{7}$$

where t_i is the commonly reported t-value, and $1/SE_i$ is the precision of an estimate. Note that the intercept and slope coefficients are reversed from the OLS version, Equation (6). Testing H_0: $\gamma_1 = 0$ (the 'precision-effect test' [PET]) from (6) or (7) provides a valid basis for determining whether there is a genuine empirical effect beyond publication selection bias (Stanley, 2008). The weakness of this linear approximation becomes apparent when one attempts to use $\hat{\gamma}_1$ as the corrected estimate of the true effect. Although PET provides a valid test for the presence of a genuine nonzero effect, $\hat{\gamma}_1$ is downwardly biased, as seen in Figure 2. The reason for this apparent discrepancy is easily explained when one realizes that the linear relation between reported effects and their SEs can be derived as a special case of the conditional mean of a truncated distribution when the underlying true effect, μ, is in fact zero.

When the underlying empirical effect is zero (i.e., $\mu = 0$), Equation (4) simplifies to $\lambda(c)$, implying that the slope of the expected effect relation reduces to this inverse Mills' ratio. Further recall that $c = a - \mu/\sigma_i$, which is merely a for $\mu = 0$. Thus, $\partial E(effect_i|truncation)/\partial \sigma_i$ reduces to the inverse Mills' ratio evaluated at the critical value of the standard normal distribution, $\lambda(a)$, which is just a constant. When there is no genuine empirical effect, the slope of expected reported effect is a constant, and the expected reported effect and its SE will be linearly related, as illustrated by line E(Effect|Selection) in Figures 1 and 2. This observation is important because it further validates the PET (H_0: $\gamma_1 = 0$). Because the null hypothesis assumes that there is no underlying effect, $\mu = 0$, a linear relation of reported effect and the SE provides a valid foundation for testing whether there is a genuine nonzero empirical effect.

To recap, the aforementioned discussion and past simulations demonstrate that a simple linear relation between an estimate and its SE may be used to test both for the presence of publication selection bias and genuine true effect beyond publication bias (Egger *et al*., 1997; Stanley, 2008). However, this linear approximation is also known to give biased estimates of the underlying true effect, μ, when $\mu \neq 0$ (Stanley, 2008). Our approach is to appeal to a higher order. In particular, we recommend using the WLS estimate of β_1 from a quadratic approximation:

$$effect_i = \beta_1 + \alpha_2 SE_i^2 + \varepsilon_i \quad \text{or} \tag{8}$$

$$t_i = \alpha_2 SE_i + \beta_1(1/SE_i) + u_i. \tag{9}$$

where MRA (8) uses WLS and $1/SE_i^2$ for weights. Note that this quadratic model of publication selection is constrained to have $\alpha_1 = 0$.

Elsewhere, $\hat{\beta}_1$ has been called the 'precision-effect estimate with SE' (PEESE) (Stanley and Doucouliagos, 2007; Havranek, 2010; Costa-Font *et al*., 2011; Doucouliagos *et al*., 2012; Stanley and Doucouliagos, 2012; Costa-Font *et al*., 2013). This MRA correction for publication bias was first proposed in Stanley (2006), refined and applied in Stanley and Doucouliagos (2007), comprehensively simulated by Moreno *et al*. (2009a), and further refined and applied in

Table 1. Means of the intercept of polynomial approximations ($n = 80$).

Heterogeneity*	True effect	Selection incidence (%)	Linear $\hat{\gamma}_1$	Quadratic	Cubic	PEESE, $\hat{\beta}_1$
	0	0	0.00	−0.01	0.00	0.00
	0	25	0.04	−0.08	0.04	0.13
	0	50	0.06	−0.11	0.07	0.25
	0	75	0.07	−0.07	0.14	0.36
$I^2 = 25\%$	0	100	0.07	0.05	0.12	0.46
	1	0	1.00	1.00	1.01	1.00
	1	25	0.92	0.94	1.08	0.99
	1	50	0.85	0.89	1.11	0.98
	1	75	0.77	0.88	1.15	0.96
	1	100	0.68	0.87	1.12	0.94
	0	0	0.00	0.00	0.00	0.00
	0	25	0.04	−0.05	−0.05	0.15
	0	50	0.08	−0.03	−0.06	0.29
	0	75	0.14	0.04	−0.02	0.44
$I^2 = 58\%$	0	100	0.20	0.18	0.07	0.60
	1	0	1.00	0.99	0.99	1.00
	1	25	0.94	0.94	1.01	1.01
	1	50	0.88	0.89	1.00	1.02
	1	75	0.81	0.85	0.91	1.02
	1	100	0.74	0.84	0.83	1.03
	0	0	0.00	0.01	0.01	0.00
	0	25	0.04	0.03	−0.09	0.18
	0	50	0.10	0.10	−0.10	0.38
	0	75	0.22	0.19	−0.09	0.61
$I^2 = 85\%$	0	100	0.37	0.34	0.09	0.86
	1	0	1.00	1.00	1.02	1.00
	1	25	0.97	0.96	0.87	1.06
	1	50	0.93	0.90	0.76	1.12
	1	75	0.87	0.86	0.62	1.17
	1	100	0.80	0.82	0.47	1.21

PEESE, precision-effect estimate with standard error.
*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. Linear, Quadratic, Cubic, and PEESE refer to different estimates of the intercept of the polynomial approximation to the conditional mean of a truncated distribution—Equation (5). In all cases, weighted least squares is used.

Doucouliagos and Stanley (2009), Moreno *et al.* (2009b, 2011), and Stanley and Doucouliagos (2012). Next, we report simulations of PEESE's bias and mean squared error (MSE) and compare them to alternative approximations and estimates, including $\hat{\gamma}_1$ from the linear approximation to this relation, which is the Egger regression.

3. Simulations

The design of our simulations closely follows Stanley (2008) and Stanley *et al.* (2010). The range of parameters employed is selected to mirror observed properties from several published meta-analyses. Briefly, random data are generated and used to test whether a regression coefficient is zero. Random heterogeneity and residuals are drawn from independent normal distributions. Regression is chosen because it is the most common statistical technique employed in the social sciences, and it encompasses many other statistical tests, including analysis of variance, *t*-tests, and tests of fixed-effects (Moore, 1997; Stanley *et al.*, 2010). Thus, the true effect (μ) in these simulations is a regression slope, and it is controlled to be either 0 or 1. Because these effects are regression slopes, they are not 'standardized' or 'effect sizes' as meta-analysts understand these terms. However, these simulations further control the magnitude of the true correlation coefficient, ρ, to be either 0 or 0.30 as the true slope coefficient is 0 or 1. When the true slope coefficient is equal to one, the effect size that is being estimated is 'small' by conventional guidelines (Cohen, 1988). Although many correlations in economics are much larger (e.g., aggregate consumer expenditures and income has a correlation of 0.999 in the USA), this small effect size, $\rho = 0.30$, is chosen to be 'conservative' and to disadvantage PEESE. Likewise, all sample sizes used to estimate this regression slope coefficient and thereby to represent the primary

Table 2. Mean square errors of precision-effect estimate with standard error (times 1000 with $n = 80$).

Heterogeneity*	True effect	Selection incidence (%)	Linear $\hat{\gamma}_1$	Quadratic	Cubic	PEESE, $\hat{\beta}_1$
	0	0	27	195	1808	8
	0	25	25	206	2148	24
	0	50	22	193	2224	68
	0	75	17	134	1749	135
$I^2 = 25\%$	0	100	10	42	422	214
	1	0	27	200	1837	8
	1	25	30	192	1783	8
	1	50	45	191	1700	8
	1	75	74	182	1627	8
	1	100	115	157	1385	10
	0	0	51	344	2862	16
	0	25	45	341	3357	35
	0	50	43	303	3103	97
	0	75	44	219	2353	204
$I^2 = 58\%$	0	100	51	115	708	359
	1	0	50	347	2927	15
	1	25	50	317	2684	15
	1	50	56	312	2529	13
	1	75	73	284	2266	12
	1	100	99	232	1841	11
	0	0	116	616	3826	37
	0	25	104	598	3952	63
	0	50	94	525	3414	168
	0	75	108	400	2466	385
$I^2 = 85\%$	0	100	172	281	955	745
	1	0	114	621	3819	36
	1	25	101	552	3414	35
	1	50	93	477	3088	41
	1	75	88	421	2716	52
	1	100	95	330	2111	65

PEESE, precision-effect estimate with standard error.
*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. Linear, Quadratic, Cubic, and PEESE refer to different estimates of the intercept of the polynomial approximation to the conditional mean of a truncated distribution—Equation (5). In all cases, weighted least squares is used.

literature estimates are quite modest for regression analysis—$n = \{30, 50, 75, 100, \text{or } 200\}$. See Stanley (2008) for more complete details.

Publication selection is modeled as the repeated sampling from these distributions until a statistically positive regression coefficient is obtained. If a given set of generated data, errors, and random heterogeneity does not produce a significant regression coefficient, an entirely new set of data, errors, and random heterogeneity is generated. This process continues until a statistically positive regression coefficient is found by chance. However, we know that not all reported scientific findings are the result of publication selection because almost all areas of research report at least a few insignificant estimates. To ensure that our simulations are realistic and robust, varying incidences of publication selection are modeled (0%, 25%, 50%, 75%, and 100%). For example, when the incidence of publication selection is 75%, exactly three fourths of the reported values have been chosen to be statistically significant, whereas the first estimate generated, significant or not, is reported for the remaining 25% of the reported values.

Here, meta-regression sample sizes are either 20 or 80. In economics, most areas of empirical research have many times more estimates. Among 87 areas of economics research, the average number of reported estimates exceeds 200 (Doucouliagos and Stanley, 2013). In medical research, there tend to be fewer randomized controlled trials (RCTs) on a given topic. But some areas of medical research have more than enough estimates. For example, Turner *et al.* (2008) reported findings on 74 antidepressant trials, and Stead *et al.* (2008) reported 42 RCTs of nicotine replacement therapy (NRT) using the 'patch' and 112 trials when other delivery systems are included. The meta-regression sample size of 20 is chosen because it is a rather small sample size for any regression estimate, whereas 80 is both practically feasible in many cases and gives these MRAs tests power to spare. Needless to say, regression-based estimators may not be appropriate if only a very small number of comparable empirical estimates exist. However, conventional fixed-effects

Table 3. Means of alternative research summary estimators ($n = 80$).

Heterogeneity*	True effect	Selection incidence (%)	Simple average	FEE	REE	*Top10*	PEESE, $\hat{\beta}_1$	PET-PEESE
	0	0	0.00	0.00	0.00	0.00	0.00	−0.01
	0	25	0.23	0.20	0.22	0.13	0.13	0.04
	0	50	0.47	0.39	0.43	0.28	0.25	0.07
	0	75	0.70	0.59	0.63	0.41	0.36	0.08
$I^2 = 25\%$	0	100	0.93	0.78	0.78	0.55	0.46	0.16
	1	0	1.00	1.00	1.00	1.00	1.00	1.00
	1	25	1.07	1.04	1.04	1.00	0.99	0.99
	1	50	1.13	1.08	1.09	1.01	0.98	0.98
	1	75	1.20	1.11	1.13	1.02	0.96	0.96
	1	100	1.26	1.15	1.16	1.02	0.94	0.94
	0	0	0.00	0.00	0.00	0.00	0.00	−0.01
	0	25	0.27	0.23	0.25	0.14	0.15	0.03
	0	50	0.54	0.45	0.51	0.30	0.29	0.09
	0	75	0.81	0.68	0.75	0.47	0.44	0.16
$I^2 = 58\%$	0	100	1.08	0.91	0.92	0.66	0.60	0.43
	1	0	1.00	1.00	1.00	1.00	1.00	1.00
	1	25	1.10	1.06	1.08	1.02	1.01	1.01
	1	50	1.19	1.13	1.16	1.04	1.02	1.02
	1	75	1.29	1.19	1.23	1.07	1.02	1.02
	1	100	1.39	1.26	1.30	1.08	1.03	1.03
	0	0	0.00	0.00	0.00	0.00	0.00	−0.02
	0	25	0.36	0.29	0.34	0.18	0.18	0.02
	0	50	0.72	0.58	0.68	0.38	0.38	0.10
	0	75	1.09	0.88	1.02	0.62	0.61	0.26
$I^2 = 85\%$	0	100	1.45	1.20	1.29	0.88	0.86	0.72
	1	0	1.00	1.00	1.00	1.00	1.00	0.98
	1	25	1.18	1.13	1.17	1.06	1.06	1.04
	1	50	1.36	1.26	1.33	1.12	1.12	1.09
	1	75	1.54	1.39	1.49	1.17	1.17	1.15
	1	100	1.73	1.52	1.63	1.22	1.21	1.20

PEESE, precision-effect estimate with standard error; PET-PEESE, precision-effect test-precision-effect estimate with standard error.

*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_\varepsilon^2)$. FEE and REE denote the fixed-effects and random-effects estimators, respectively. *Top10* is the simple average of the most precise 10% of the observations. $\hat{\beta}_1$ is estimated from Equation (9).

and random-effects meta-analysis suffer from *nearly* the same limitation, because they too are WLS regression estimates (Raudenbush, 1994; Stanley and Doucouliagos, 2013). Although there is no generally applicable lower limit for the sample size of a MA or a MRA, PEESE will require one more observation than the conventional fixed-effects meta-analysis because it must estimate two parameters compared with fixed-effect's one.

In addition to the incidence of publication selection, the statistical properties of these alternative estimators are most influenced by the relative magnitude of the unexplained heterogeneity relative to the sampling errors. We use Higgins and Thompson's (2002) $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_\varepsilon^2)$ as the indicator of the size of the relative heterogeneity. σ_h^2 is the between-study heterogeneity variance, and σ_ε^2 is the within-study sampling variance. I^2 is analogous to R^2 in regression analysis. It reflects the proportion of the total variation due to unexplained heterogeneity. Simulations are conducted over a wide range of heterogeneity and publication selection and reported in Tables 1–4 for $n = 80$, whereas the simulation results for $n = 20$ are reported in Tables 5–8. Although the exact calculated value of I^2 varies for each random sample, these tables state its population value when there is no publication selection.

Table 1 reports the average of 10,000 replications for alternative polynomial approximations, Equation (5), and Table 2 the associated MSE of these approximations. Recall that true effects (μ) are either 0 or 1. In all cases, the estimated intercept,β_1, is used as the corrected estimate in a WLS version of Equation (5). The first column of simulation results reports the 'linear' approximation (i.e., K = 1) of Equation (5), which is equivalent to $\hat{\gamma}_1$ from Equation (7). Next is the 'quadratic' approximation (i.e., K = 2), followed by the 'cubic' approximation (i.e., K = 3). Lastly is our recommended PEESE estimator, which is the quadratic approximation that further constrains

Table 4. Mean square errors of alternative research summary estimators (times 1000 with $n = 80$).

Heterogeneity*	True effect	Selection incidence (%)	Simple average	FEE	REE	*Top10*	PEESE, $\hat{\beta}_1$	PET-PEESE
	0	0	3	3	3	14	8	22
	0	25	58	41	49	33	24	23
	0	50	221	156	187	87	69	23
	0	75	494	344	396	180	135	24
$I^2 = 25\%$	0	100	875	603	603	310	214	73
	1	0	3	3	3	14	8	8
	1	25	7	4	5	13	8	8
	1	50	20	8	10	13	8	8
	1	75	41	15	18	13	8	8
	1	100	71	25	28	13	10	10
	0	0	6	6	5	27	16	42
	0	25	78	55	69	49	35	42
	0	50	295	207	260	115	97	45
	0	75	658	464	560	243	203	65
$I^2 = 58\%$	0	100	1168	829	839	447	358	255
	1	0	6	6	5	27	15	16
	1	25	15	9	11	26	15	16
	1	50	42	22	29	25	13	14
	1	75	88	42	58	25	12	13
	1	100	152	70	93	26	11	11
	0	0	16	14	14	64	36	96
	0	25	145	94	129	97	63	96
	0	50	535	344	477	206	170	95
	0	75	1087	788	1041	422	388	159
$I^2 = 85\%$	0	100	2100	1435	1672	792	743	619
	1	0	16	15	12	63	36	59
	1	25	46	30	39	59	35	58
	1	50	144	80	120	63	41	64
	1	75	305	162	245	71	52	72
	1	100	534	273	406	82	64	74

PEESE, precision-effect estimate with standard error; PET-PEESE, precision-effect test-precision-effect estimate with standard error.

*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. FEE and REE denote the fixed-effects and random-effects estimators, respectively. *Top10* is the simple average of the most precise 10% of the observations. $\hat{\beta}_1$ is estimated from Equation (9).

$\alpha_1 = 0$. PEESE is the same as estimating $\hat{\beta}_1$ in Equation (9). With 10,000 replications, the results are stable. For example, the average $\hat{\beta}_1$ varies only in the fourth decimal place, and its coefficient of variation is less than 0.06% when these 10,000 replications are repeated an additional 10 times.

Although the shape of bias (Table 1) is rather complex, a few clear patterns emerge, especially when one considers both bias and efficiency as measured by MSE. First, PEESE ($\hat{\beta}_1$) has the smallest MSE in the great majority (70%) of cases, often by a wide margin (Table 2), and it also has the smallest bias in a plurality of simulations. However, PEESE is upwardly biased when the true effect is zero. Second, the PET coefficient, $\hat{\gamma}_1$, dominates PEESE as expected when $\mu = 0$. Recall that the linear approximation is correctly specified when the true effect is zero. Nonetheless, in a few incidences, either the quadratic or the cubic approximation has a smaller bias than $\hat{\gamma}_1$. Like PEESE, it is easy to see that $\hat{\gamma}_1$ is upwardly biased when $\mu = 0$. Perhaps, this upward bias is a reflection of attenuation bias (or, equivalently, 'errors-in-variables' bias) that will result from using a fallible estimate, SE_i, in the place of σ_i. Third, the unconstrained quadratic and cubic approximations are clearly inferior to either $\hat{\beta}_1$ or $\hat{\gamma}_1$. Their MSEs are typically many times larger than these other approximations. The few cases where they have a slightly smaller bias seem random and unpredictable unless we were to know the exact incidence of publication selection. In practice, we have no way to know the percent of estimates that have been selected.

The unreliability of the unconstrained quadratic and cubic approximations is likely caused by multicollinearity among powers of SE. These powers of SE are highly correlated. For example, the variance inflation factor for the unconstrained quadratic approximation using the data shown in Figure 1 is 6.5 and 166 for the cubic

Table 5. Means of the intercept of polynomial approximations ($n = 20$).

Heterogeneity*	True effect	Selection incidence (%)	Linear $\hat{\gamma}_1$	Quadratic	Cubic	PEESE, $\hat{\beta}_1$
	0	0	0.00	0.01	−0.01	0.00
	0	25	0.04	−0.09	0.01	0.13
	0	50	0.06	−0.10	0.15	0.25
	0	75	0.07	−0.08	0.19	0.36
$I^2 = 25\%$	0	100	0.07	0.07	0.15	0.46
	1	0	1.00	1.01	1.01	1.00
	1	25	0.92	0.97	1.14	0.99
	1	50	0.86	0.91	1.14	0.98
	1	75	0.77	0.89	1.18	0.96
	1	100	0.69	0.88	1.10	0.94
	0	0	0.00	0.02	0.01	0.00
	0	25	0.05	−0.06	0.02	0.15
	0	50	0.09	−0.06	−0.13	0.30
	0	75	0.14	0.04	−0.03	0.44
$I^2 = 58\%$	0	100	0.20	0.19	0.09	0.59
	1	0	1.00	1.00	1.08	1.00
	1	25	0.94	0.93	1.03	1.01
	1	50	0.88	0.87	0.98	1.02
	1	75	0.82	0.87	0.95	1.03
	1	100	0.74	0.86	0.79	1.03
	0	0	0.00	0.01	0.04	0.00
	0	25	0.02	0.03	−0.10	0.17
	0	50	0.10	0.08	−0.11	0.38
	0	75	0.23	0.20	−0.19	0.61
$I^2 = 85\%$	0	100	0.39	0.36	0.07	0.86
	1	0	1.00	0.98	0.96	1.00
	1	25	0.98	0.99	0.90	1.07
	1	50	0.93	0.92	0.82	1.12
	1	75	0.91	0.88	0.63	1.19
	1	100	0.82	0.84	0.42	1.22

PEESE, precision-effect estimate with standard error.
*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. Linear, Quadratic, Cubic, and PEESE refer to different estimates of the intercept of the polynomial approximation to the conditional mean of a truncated distribution—Equation (5). In all cases, weighted least squares is used.

approximation. This multicollinearity-induced unreliability is clearly seen in the large MSEs of the cubic model (Table 2). The MSEs of the cubic model get much worse still for $n = 20$ (Table 6). Our constrained quadratic, Equation (8), as well as the linear approximation, has no multicollinearity; hence, the resulting estimators are much more efficient.

Several implications and suggestions can be drawn from the relative bias and efficiency of these alternative approximations. First, both unconstrained polynomial approximations are distinctly inferior and can thereby be eliminated from further consideration. Second, PEESE dominates the linear approximation, $\hat{\gamma}_1$, when there is a genuine nonzero effect. Third, the opposite is largely true when there is no genuine effect. This suggests that a combined estimator may be better than either PEESE or $\hat{\gamma}_1$, by themselves. We propose that the PEESE estimator be used only when there is evidence of a nonzero effect (reject H_0: $\hat{\gamma}_1 = 0$) in Equation (7). When PET is not passed (i.e., accept H_0: $\gamma_1 = 0$), $\hat{\gamma}_1$ should be used as the corrected estimate. We call this conditional estimator, 'PET-PEESE,' and its bias and MSE are reported in Tables 3 and 4 along with alternative conventional meta-analysis summary estimates.

Tables 3 and 4 display the bias and efficiency of PEESE, the combined estimator, PET-PEESE, and several conventional summary meta-estimates. The fixed-effects and random-effects estimators (FEE and REE) are weighted averages of the reported effects, where the weights are the inverse of the estimates' variances. REE employs a more complex variance estimate that includes the between-study variance, σ_h^2 (Cooper and Hedges, 1994). In our simulations, excess unexplained heterogeneity is always included; thus, by conventional practice, REE should be preferred over FEE. However, conventional practice is wrong when there is publication selection. With selection for statistical significance, REE is always more biased than FEE (Table 3). This predictable inferiority is due to the fact that REE is itself a weighted average of the simple mean, which has the largest publication bias, and FEE. Both weighted averages are less biased than

Table 6. Mean square errors of polynomial approximations (times 1000 with $n = 20$).

Heterogeneity*	True effect	Selection incidence (%)	Linear $\hat{\gamma}_1$	Quadratic	Cubic	PEESE, $\hat{\beta}_1$
	0	0	105	910	12,805	31
	0	25	93	960	15,507	45
	0	50	76	837	15,772	84
	0	75	55	629	13,268	143
$I^2 = 25\%$	0	100	25	196	2889	217
	1	0	105	960	12,749	32
	1	25	108	909	12,117	31
	1	50	114	861	11,916	29
	1	75	138	778	10,694	28
	1	100	177	677	9236	28
	0	0	208	1705	21,298	64
	0	25	182	1719	25,583	78
	0	50	154	1506	24,863	133
	0	75	126	1082	18,969	231
$I^2 = 58\%$	0	100	89	452	5468	362
	1	0	208	1690	21,122	64
	1	25	196	1553	19,738	60
	1	50	186	1432	19,300	55
	1	75	182	1242	16,471	48
	1	100	193	1015	13,080	41
	0	0	482	3350	33,902	150
	0	25	447	3429	37,066	167
	0	50	377	3074	34,960	256
	0	75	316	2190	24,800	454
$I^2 = 85\%$	0	100	281	1033	8735	778
	1	0	494	3416	34,339	153
	1	25	418	3014	32,328	134
	1	50	373	2553	28,468	131
	1	75	320	2196	22,638	133
	1	100	270	1645	17,405	124

PEESE, precision-effect estimate with standard error.
*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. Linear, Quadratic, Cubic, and PEESE refer to different estimates of the intercept of the polynomial approximation to the conditional mean of a truncated distribution—Equation (5). In all cases, weighted least squares is used.

the simple mean because they give greater weight to the less selected and thereby less biased estimates, which also tend to be the most precise (recall the discussion in Section 2).

The simple average is included in Table 3 and 4 to document how large the publication biases are, when there is selective reporting of scientific results. The magnitude of this bias can be especially severe when there is no genuine underlying empirical effect. *Top10* is a more radical weighted average introduced by Stanley *et al*. (2010) to emphasize the importance of publication bias for scientific inference. *Top10* is the simple average of the most precise 10% (smallest SEs) of the reported research results. That is, 90% of research results are assigned a weight of 0, whereas the most precise 10% are given a weight of 1. Publication bias is such a serious threat to the integrity of scientific inference that it is often better to just throw out 90% of the reported research (Stanley *et al*., 2010). For all incidences of selection, *Top10* has smaller bias than any of the conventional summary statistics that use all the research results. Throwing away 90% of the research is more efficient in the majority of cases (Table 4). Nonetheless, the meta-regression estimators derived here are clearly better than the *Top10* and the more conventional summary statistics.

We do not report the statistical properties for the widely used nonparametric 'Trim & Fill' correction strategy (Duval and Tweedie, 2000), because previous simulations have shown that both PEESE and the PET coefficient, $\hat{\gamma}_1$: 'consistently outperformed the Trim & Fill estimators. . . . With respect to the popular Trim & Fill method, we find it hard to recommend over the regression-based alternatives due to its potentially misleading adjustments and poor coverage probabilities' (Moreno *et al*., 2009a, p. 1 and 12).

For ease of comparison, we report the simulation results for PEESE ($\hat{\beta}_1$) in Tables 3 and 4 along with our new hybrid estimator, PET-PEESE. First, notice how PEESE dominates all of the conventional summary estimators and *Top10*. Table 4 shows very clearly that PEESE has smaller MSE when there is publication selection. Even when there is no selection, PEESE has only slightly larger variance. Otherwise, there is little reason to use any of the better

Table 7. Means of alternative research summary estimators ($n = 20$).

Heterogeneity*	True effect	Selection incidence (%)	Simple average	FEE	REE	*Top10*	PEESE, $\hat{\beta}_1$	PET-PEESE
	0	0	0.00	0.00	0.00	0.00	0.00	−0.02
	0	25	0.23	0.20	0.21	0.14	0.13	0.03
	0	50	0.47	0.39	0.43	0.28	0.25	0.06
	0	75%	0.70	0.59	0.63	0.41	0.36	0.07
$I^2 = 25\%$	0	100	0.93	0.78	0.78	0.55	0.46	0.10
	1	0	1.00	1.00	1.00	1.00	1.00	0.98
	1	25	1.07	1.04	1.05	1.01	0.99	0.95
	1	50	1.13	1.08	1.09	1.01	0.98	0.93
	1	75	1.20	1.11	1.13	1.02	0.96	0.90
	1	100	1.27	1.15	1.17	1.03	0.94	0.86
	0	0	0.00	0.00	0.00	0.00	0.00	−0.03
	0	25	0.27	0.23	0.26	0.15	0.15	0.03
	0	50	0.54	0.45	0.51	0.31	0.30	0.09
	0	75	0.81	0.68	0.75	0.48	0.44	0.15
$I^2 = 58\%$	0	100	1.08	0.91	0.92	0.67	0.59	0.27
	1	0	1.00	1.00	1.00	1.00	1.00	0.93
	1	25	1.09	1.07	1.08	1.03	1.01	0.93
	1	50	1.19	1.13	1.16	1.05	1.02	0.92
	1	75	1.29	1.19	1.23	1.07	1.03	0.91
	1	100	1.39	1.26	1.30	1.09	1.03	0.90
	0	0	0.00	0.00	0.00	0.00	0.00	−0.04
	0	25	0.36	0.28	0.34	0.19	0.17	0.00
	0	50	0.72	0.58	0.68	0.39	0.38	0.10
	0	75	1.09	0.89	1.02	0.64	0.61	0.23
$I^2 = 85\%$	0	100	1.44	1.20	1.29	0.90	0.86	0.47
	1	0	1.00	1.00	1.00	1.00	1.00	0.88
	1	25	1.18	1.13	1.17	1.07	1.07	0.91
	1	50	1.36	1.26	1.33	1.14	1.12	0.92
	1	75	1.54	1.40	1.49	1.19	1.19	0.95
	1	100	1.72	1.52	1.63	1.23	1.22	0.96

PEESE, precision-effect estimate with standard error; PET-PEESE, precision-effect test-precision-effect estimate with standard error.

*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_e^2)$. FEE and REE denote the fixed-effects and random-effects estimators, respectively. *Top10* is the simple average of the most precise 10% of the observations. $\hat{\beta}_1$ is estimated from Equation (9).

known summary statistics in a systematic review. Only *Top10* has smaller bias in any of these simulation combinations, and this occurs only in a small minority of cases.

Lastly, note that our conditional estimator ($\hat{\beta}_1$ when we reject H_0: $\gamma_1 = 0$ and $\hat{\gamma}_1$ when we fail to reject it) improves upon both PEESE and the PET coefficient, $\hat{\gamma}_1$. PET-PEESE has a smaller MSE than $\hat{\gamma}_1$ in 70% of the cases, and it has a smaller bias than PEESE in the majority of cases. On the other hand, PEESE has smaller bias than PET-PEESE in only 17% of the cases (Table 3). If there is any selection for statistical significance, PET-PEESE has equal or smaller bias, in some cases by several times. When there is no publication selection the conditional estimator has a very small downward bias. Overall, however, PET-PEESE has the smallest average bias among any of these estimators. When it comes to efficiency, the simulations are less favorable to PET-PEESE. Nonetheless, it has equal or smaller MSE than PEESE in the majority of cases, and recall that PEESE is more efficient than any of these other estimates in the great majority of cases (Table 4). Thus, our new conditional estimator is the best choice whenever a research literature is suspected to contain publication selection, and such a suspicion will be warranted for most empirical literatures across the social, medical, and natural sciences.

These meta-regression methods do not perform quite as strongly when there are only 20 estimates available ($n = 20$)—see Tables 5–8. Nonetheless, they still have lower average bias and MSE than the conventional alternatives. Even when there are only 20 estimates, PEESE has the lowest average MSE, and PET-PEESE has the lowest average bias.

In spite of these favorable findings, we would be remiss if we did not recommend some caution. The largest threat to these meta-regression methods of publication bias reduction occurs when there is no

Table 8. Mean square errors of alternative estimators (times 1000 with $n = 20$).

Heterogeneity*	True effect	Selection incidence (%)	Simple average	FEE	REE	*Top10*	PEESE, $\hat{\beta}_1$	PET-PEESE
	0	0	13	11	11	56	31	89
	0	25	65	47	55	83	44	86
	0	50	228	161	191	135	84	74
	0	75	497	347	397	214	144	57
$I^2 = 25\%$	0	100	878	606	608	323	217	53
	1	0	13	11	11	55	32	53
	1	25	16	12	12	54	31	63
	1	50	27	15	17	51	29	69
	1	75	47	22	25	50	28	85
	1	100	77	31	35	49	28	103
	0	0	24	22	20	112	64	173
	0	25	92	69	81	143	78	161
	0	50	307	220	270	202	134	144
	0	75	667	474	566	307	231	128
$I^2 = 58\%$	0	100	1171	834	858	476	368	167
	1	0	23	23	21	112	64	132
	1	25	30	25	26	103	60	139
	1	50	54	35	41	97	55	147
	1	75	97	53	68	92	48	150
	1	100	161	81	103	87	41	155
	0	0	62	58	56	270	150	400
	0	25	182	130	162	336	167	393
	0	50	564	376	500	428	256	343
	0	75	1211	819	1060	614	454	307
$I^2 = 85\%$	0	100	2113	1461	1692	906	778	425
	1	0	63	58	55	284	153	322
	1	25	87	68	76	257	134	304
	1	50	178	113	150	237	131	306
	1	75	330	195	273	218	133	297
	1	100	554	298	427	208	124	291

PEESE, precision-effect estimate with standard error; PET-PEESE, precision-effect test-precision-effect estimate with standard error.

*Heterogeneity is measured by $I^2 = \sigma_h^2/(\sigma_h^2 + \sigma_\varepsilon^2)$. FEE and REE denote the fixed-effects and random-effects estimators, respectively. *Top10* is the simple average of the most precise 10% of the observations. $\hat{\beta}_1$ is estimated from Equation (9).

genuine underlying empirical effect (i.e., $\mu = 0$). In these cases, all estimators are biased if there is selection for statistical significance. In the unlikely case that all studies are prepared to report only significantly positive effects, very large biases are manufactured. However, even under such worse case scenarios, PET-PEESE has a much smaller bias than the other alternatives, reducing the publication bias seen in the simple mean by at least half and often much more. When there is evidence of publication bias (reject H_0: $\alpha_1 = 0$ in Equation (7)) but no evidence of an underlying empirical effect (accept H_0: $\gamma_1 = 0$), caution might suggest that we offer no summary estimate of effect. Second, meta-regression methods (and *Top10)* are unlikely to be reliable when there are only a few comparable research results in a given area. In very small samples, FEE is likely to provide the best summary of a systematic review, because it will be less biased than REE. However, FEE tends to give confidence intervals that are too small, but this can be easily remedied through WLS (Stanley and Doucouliagos, 2013). Nonetheless, when there are as few as 20 estimates, these meta-regression approximations still fare rather well relative to alternative methods.

In sum, when there are sufficient reported estimates, we advocate that meta-analysts first run MRA (7). If they find evidence of a genuine empirical effect (reject H_0: $\gamma_1 = 0$), then use $\hat{\beta}_1$ from MRA (9) as the corrected estimate of effect. Otherwise, $\hat{\gamma}_1$ should be employed. Of course, MRA models (6) and (8) may be used in place of Equations (7) and (9), respectively, when a WLS statistical routine is also employed. To be conservative, one should always use either $\hat{\beta}_1$ or $\hat{\gamma}_1$ even if there is insufficient evidence of publication selection (i.e., accept H_0: $\alpha_1 = 0$ in Equation (7)) because the Egger test is known to have low power (Egger *et al*., 1997; Stanley, 2008).

4. Examples

In many areas of empirical science, correcting for publication bias will make an important practical difference to our understanding. For example, the magnitude of the value of a statistical life (VSL) is a critical parameter for many public health and safety initiatives. The VSL measures the trade-off between money and risk for very small risks of death as revealed by the choices that workers and consumers are observed to make. These statistical estimates may be derived from a regression of workers' wages on the risk of a job-related fatality along with other determinants of wages (Viscusi, 1993). A meta-analysis of 39 separate VSL estimates finds the average VSL to be $9.5 million (Bellavance *et al*., 2009). Figure 3 plots these 39 estimates of VSL, measured in millions of US dollars and corrected for inflation.

Column 1 of Table 9 reports the MRA findings for these VSL estimates using meta-regression models (6) and (8) using (WLS) and $1/SE_i^2$ for the weights. The VSL is reduced by 82% when publication selection is considered; PEESE = $1.67 million—see column 1 Table 9 and Doucouliagos *et al*. (2012). Needless to say, there is clear evidence of publication bias (reject H_0:$\alpha_1 = 0$; $p < 0.01$), and this is reflected by the highly skewed funnel graph —Figure 3. Which researcher would be willing to report that the value of life is negative? Furthermore, there is strong evidence that VSL is genuinely larger than zero (reject H_0:$\gamma_1 = 0$; $p < 0.01$); thus, PET-PEESE would also be $1.67 million. Needless to say, reducing VSL by 82% greatly affects the number of health and safety projects and regulations that are socially beneficial (or cost effective).

The adverse employment effect from a rise in the minimum wage is another important dimension for public policy. Raising the minimum wage always engenders a public controversy that is often stated in terms of harm to workers. When we apply these methods to 1474 elasticity estimates of the effect of minimum wage on employment, a small adverse employment effect, −0.19, is reduced to one that is both statistically and practically insignificant, −0.009—see column 2 of Table 9 (Doucouliagos and Stanley, 2009). These effects are measured in terms of elasticity, which, in this case, measures the percent decrease in teen employment that results from a 1% increase in the minimum wage. Because we accept H_0:$\gamma_1 = 0$, $\hat{\gamma}_1 = -0.009$ is our preferred estimate. Our corrected estimate of effect, −0.009, implies that a doubling of the minimum wage would cause a less than 1% reduction of teen employment.

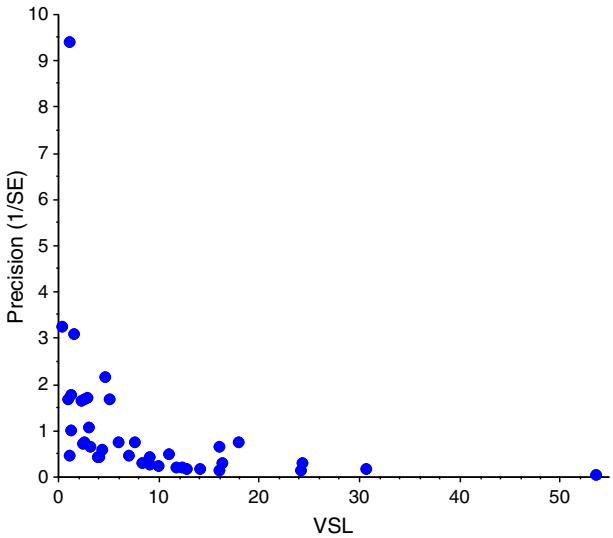

Figure 3. Funnel plot of the estimated value of a statistical life (in millions of US dollars) with their precisions (1/SE) on the vertical axis.

Table 9. Corrected estimates and weighted least squares meta-regression.

Variable	1: Statistical life	2: Minimum wage	3: NRT patch	4: Antidepressants
$\hat{\alpha}_1$	3.20 (6.67)	−1.60(−17.36)	1.09 (2.38)	1.84 (5.47)
$\hat{\gamma}_1$	0.81 (3.56)	−0.0009 (−1.09)	0.197 (2.00)	0.13 (2.50)
Simple mean	$9.5 mil	−0.19	0.657	0.47
PEESE, $\hat{\beta}_1$	$1.67 mil	−0.036	0.314	0.29
n	39	1474	42	50

t-values are reported in parenthesis. $\hat{\alpha}_1$ and $\hat{\gamma}_1$ are estimated from Equation (6), and $\hat{\beta}_1$ is estimated from Equation (8).
NRT, nicotine replacement therapy; PEESE, precision-effect estimate with standard error.

No doubt, some economists may be skeptical about such a large correction of minimum wage's adverse employment effect. However, we find that a negligible practical effect from minimum wage is a very robust summary of this extensive empirical literature. This employment effect remains practically insignificant whether one uses PEESE = −0.036, *Top10* = −0.0217, or multiple MRA results that use dozens of moderator variables (Doucouliagos and Stanley, 2009). In actual applications, the simple meta-regression models of publication selection bias advanced here need to be embedded within more complex, multiple MRAs that also account for observed systematic heterogeneity. Current space does not permit a detailed discussion of the conventional econometric practice of using moderator variables in multiple meta-regression models to explain much of the observed variation among research results. See Doucouliagos and Stanley (2009); Havranek (2010); Feld and Heckemeyer (2011), and Doucouliagos *et al.* (2012) for examples. Conservatively, the modest average adverse employment effect found in the minimum-wage literature is reduced by a factor of five when observed publication selection is accommodated.

Or, for a medical example with public health policy implications, consider Stead *et al.* (2008) systematic review of all of the clinical trials of NRT for smoking cessation, 42 of which involve the 'patch.' Column 3 of Table 9 reports the MRA findings for these clinical trials measured by log risk ratios and indicates publication selection (reject H_0:$\alpha_1 = 0$; $p < 0.05$). The average log risk ratio is 0.657, which implies that smokers who use the 'patch' are 93% more likely to quit smoking. Because these clinical trials do not pass the PET (i.e., accept H_0: $\gamma_1 = 0$; $p > 0.05$), the PET coefficient, $\hat{\gamma}_1 = 0.197$, is our preferred corrected estimate. Such a correction reduces the efficacy of the patch to only 22%. The Table A1 provides this data on NRT using the 'patch' along with the SPSS and STATA commands needed to reproduce our PET-PEESE results.

Lastly, we apply these meta-regression methods to the controversial issue of the efficacy of antidepressants. Turner *et al.* (2008) collected data on all of the phase II and phase III trials of antidepressants registered at the US Department of Food and Drug Administration (FDA) and those that were also published. To sell pharmaceuticals in the USA, RCTs of their safety and efficacy must be reported to the FDA. Thus, the FDA registry of clinical trials is considered the 'gold standard.' Of these 74 RCTs of antidepressants, only 50 are published in journals (Turner *et al.*, 2008). Furthermore, Moreno *et al.* (2009b) found that PEESE is the best method to correct these 50 published RCT results for publication bias.

Figure 4 plots the funnel for the FDA gold standard using the data of Turner *et al.* (2008) with the effect size, measured by Glass's g, on the horizontal axis. In Figure 4, published trails are shown twice—first, as they were reported to the FDA ('diamond') and second, as published ('half-moon'). It is difficult to imagine a clearer depiction of selective reporting. Many published trials report different effect sizes using alternative measures from what are found in the FDA registry for the same clinical trials. This is shown in Figure 4 for those 'half-moons' that are clear on their left sides. If the same results were reported to the FDA as those published, then a diamond would show through the open left side of that half-moon. Chan *et al.* (2004) found that published clinical trials often report different outcome measures than those stated in their research protocols. Needless to say, the funnel-asymmetry test (FAT; H_0:$\alpha_1 = 0$; Equation (6) finds significant publication selection for positive effects ($t = 5.47$; $p < 0.01$; column 4 of Table 9).

Fortunately, there is also evidence of a genuine positive clinical effect from taking antidepressants ($t = 2.50$; $p < 0.05$). However, the modest average effect size of 0.47 is exaggerated by over 60% when compared with PEESE = 0.29. PEESE-MRA model (8) is represented by the curve in Figure 4. Note how $\hat{\beta}_1$ is twice as large as

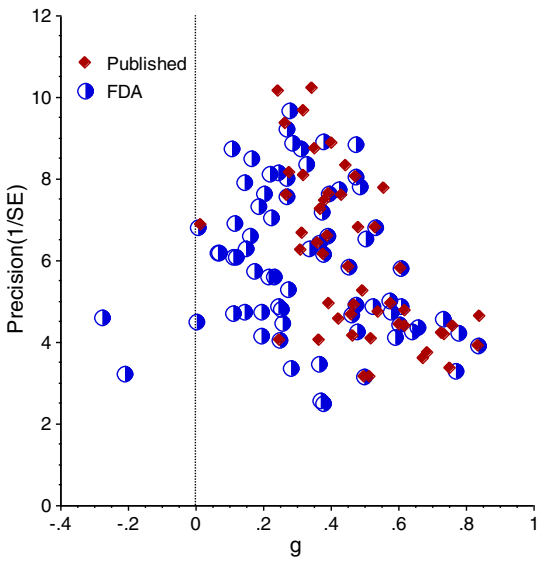

Figure 4. Funnel plot of the Food and Drug Administration registry of antidepressant trials measured by Glass's *g* on the horizontal axis with their precisions (1/SE) on the vertical axis. Results that were published are depicted by a half moon, whereas diamonds are used to display both published and unpublished effect sizes of antidepressant trials.

$\hat{\gamma}_1$ here and also for our other example where PET is passed (column 1). Thus, using the appropriate approximation can make an important practical difference. Our corrected meta-regression estimate for the effect size of antidepressants is almost exactly equal to the weighted average, 0.31 (FEE and REE), of those 74 trials reported to the FDA (Moreno *et al*., 2009b; Turner *et al*., 2008). Knowing that antidepressants have a smaller effect than what published RCTs report might change clinical practice and thereby affect millions of patients.

Perhaps the biggest advantage of these meta-regression approximations for publication selection is that they easily accommodate systematic heterogeneity, publication selection (or small-study) bias, and a differential publication selection that is associated with other moderator variables. Nearly, all areas of empirical research contain excess systematic heterogeneity. That is, empirical effects depend on the population being treated, the severity of the subjects' prior conditions, dosage, the exact treatment protocol, and so on. Among hundreds of MRAs of economics research, none have found the absence of excess heterogeneity as measured by the conventional Cochran's Q-test (Cooper and Hedges, 1994). In all cases, meta-regression analysts have found that the choice of variables, econometric model, and methods makes a large practical difference to reported research results. Thus, statistically valid meta-analyses must also accommodate systematic heterogeneity. Other publication correction strategies do not (Moreno *et al*., 2009a).

Explaining reported research variation can easily be accomplished by expanding any of our meta-regression models of publication selection. For example, meta-regression model (8) becomes

$$effect_i = \beta_1 + \sum_k \delta_k Z_k + \alpha_0 SE_i^2 + \sum_j \alpha_j K_j SE_i^2 + \varepsilon_i. \quad (10)$$

where Z_k is a moderator variable that can help to explain genuine systematic variation among reported findings, and K_j is a selection variable that is related to a greater or a lesser intensity to report statistically significant positive findings. This model can address possible interaction of funnel plot asymmetry and moderator variables by simultaneously fitting a meta-regression and a publication bias model.

5. Conclusion

Publication selection bias is a widely recognized threat to the validity of empirical scientific inquiry. This threat is often so severe that a balanced assessment of the efficacy of medical treatments is difficult or impossible. This threat remains even when there have been clear findings reported from the 'gold standard' of empirical science—double-blind, placebo-controlled randomized clinical trials. In the social sciences where empirical inquiry often uses observational data, this bias is routinely much worse still (Doucouliagos and Stanley, 2013). Fortunately, there is a long history of statistical interest in this problem. Unfortunately, corrections for publication selection bias have not been widely adopted, and their performance and reliability has been wanting.

In this paper, we derive meta-regression methods that are easy to apply, are likely to greatly reduce publication selection bias in most applications and offer a practical solution to this important threat to modern science. As a side effect of investigating the theoretical foundation for our meta-regression model of publication selection, we are able to explain both the success and the bias of the Egger meta-regression model, which may be seen as a linear approximation to a complex nonlinear function. Nonetheless, this linear approximation provides practical tests of both the existence of selection and the presence of a genuine nonzero empirical effect beyond publication bias (Stanley, 2008). Unfortunately, the linear approximation does not offer a suitable corrected estimate when there is nonzero 'true' effect.

For these cases, we demonstrate how a constrained quadratic approximation, PEESE, to the conditional expected value of a truncated distribution is considerably less biased and often more efficient. Furthermore, simulations demonstrate how a hybrid between these two approximations improves the correction for publication selection bias yet further. Both approximations are very simple to apply, merely OLS of common statistics (t-values, SEs, and precision) or, equivalently, WLS of reported effects, their SEs, and/or variances.

Needless to say, these methods have limitations. First, being based on regression analysis, they require more than a few estimates on the same empirical phenomenon. Second, overwhelming unexplained heterogeneity can invalidate the underlying meta-regression tests (i.e., the PET) (Stanley, 2008). However, when unexplained heterogeneity is responsible for more than 90% of the observed variation among reported research results, uncorrected publication biases will expand greatly. Thus, balanced scientific assessment does not have the luxury to do nothing. Even in these extreme cases, the methods advanced here will remain a marked improvement over conventional meta-analytic summary statistics.

Appendix

The purpose of this appendix is to supply the reader with a dataset and the statistical commands that will enable the replication of the FAT-PET-PEESE results reported in this paper. Table A1 reports the 42 trials of NRT using the 'patch' as found in Stead *et al*. (2008). The findings reported in column 3 Table 9 can be obtained using SPSS by

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT t
  /METHOD=ENTER Precision.
```

or

```
REGRESSION
  /MISSING LISTWISE
  /REGWGT=Precision_sq
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT logRR
  /METHOD=ENTER SE.
```

The SPSS PEESE results can be found by

```
REGRESSION
  /MISSING LISTWISE
  /REGWGT=Precision_sq
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT logRR
  /METHOD=ENTER Variance.
```

These same findings will result in STATA using: '**regress logRR SE [aweight = Precision_sq]**' or '**regress t Precision**' and '**regress logRR Variance [aweight = Precision_sq]**' for PEESE.

Table A1. Randomized controlled trials of the 'patch' for nicotine replacement therapy

Study ID	Log(RR)	SE	Precision	*t*	Variance	Precision_sq
Hughes 1999	0.517974	0.173580	5.761032	2.984065	0.030130	33.189494
Paoletti 1996	1.321756	0.532291	1.878672	2.483145	0.283334	3.529407
Campbell 1996	0.379032	0.288775	3.462904	1.312551	0.083391	11.991702
Ehrsam 1991	1.252763	0.779194	1.283377	1.607768	0.607143	1.647058
Wong 1999	1.212758	0.530483	1.885075	2.286139	0.281412	3.553506
Otero 2006	0.467009	0.100209	9.979144	4.660350	0.010042	99.583307
Lewis 1998	0.530899	0.534102	1.872302	0.994003	0.285265	3.505513
CEASE 1999	0.355630	0.121645	8.220642	2.923507	0.014798	67.578955
Richmond 1994	0.721681	0.304625	3.282725	2.369080	0.092796	10.776281
TNSG 1991	0.591664	0.188984	5.291453	3.130762	0.035715	27.999477
Oncken 2007	0.123060	0.245511	4.073137	0.501240	0.060276	16.590447
Sonderskov 1997	0.418471	0.337193	2.965661	1.241043	0.113699	8.795143
Perng 1998	1.163151	0.616329	1.622510	1.887224	0.379861	2.632539
Westman 1993	2.104759	0.732925	1.364396	2.871725	0.537179	1.861577
Tonnesen 2000	1.551034	0.769624	1.299336	2.015314	0.592321	1.688273
ICRF 1994	0.362814	0.172201	5.807167	2.106922	0.029653	33.723191

Continues

Table A1. Continued

Study ID	Log(RR)	SE	Precision	*t*	Variance	Precision_sq
Hurt 1990	0.287682	0.476603	2.098182	0.603609	0.227150	4.402369
Wisborg 2000	0.070068	0.303622	3.293569	0.230774	0.092186	10.847597
Moolchan 2005	1.666596	0.746123	1.340262	2.233675	0.556700	1.796301
Fiore 1994B	0.880573	0.560602	1.783797	1.570763	0.314275	3.181931
Daughton 1991	1.233715	0.507001	1.972383	2.433358	0.257050	3.890294
Hays 1999	0.807430	0.287888	3.473573	2.804667	0.082880	12.065710
Jorenby 1999	0.558835	0.377398	2.649723	1.480758	0.142429	7.021030
Sachs 1993	0.975060	0.342810	2.917068	2.844316	0.117519	8.509284
Ahluwalia 1998	0.377294	0.245931	4.066181	1.534146	0.060482	16.533829
Tonnesen 1991	1.379374	0.441013	2.267507	3.127740	0.194492	5.141587
Abelin 1989	0.425268	0.360044	2.777438	1.181156	0.129632	7.714164
Kornitzer 1995	−0.051293	0.364186	2.745850	−0.140843	0.132631	7.539690
Killen 1997	0.090972	0.285768	3.499342	0.318342	0.081663	12.245395
Buchkremer 1988	0.376273	0.344040	2.906639	1.093690	0.118364	8.448549
Davidson 1998	0.723919	0.296337	3.374536	2.442891	0.087816	11.387496
Glavas 2003a	0.367725	0.390282	2.562250	0.942203	0.152320	6.565124
Hurt 1994	0.663294	0.269184	3.714931	2.464091	0.072460	13.800713
Prapavessis 2006	0.668636	0.432090	2.314333	1.547446	0.186702	5.356136
Velicer 2006	−0.239248	0.224243	4.459448	−1.066914	0.050285	19.886677
Cinciripini 1996	0.538997	0.404587	2.471656	1.332215	0.163691	6.109085
Daughton 1998	0.451707	0.302754	3.303012	1.491993	0.091660	10.909886
Joseph 1996	−0.201751	0.237075	4.218074	−0.851001	0.056205	17.792152
Fiore 1994A	0.510826	0.363763	2.749043	1.404282	0.132324	7.557235
Hughes 2003	0.363618	0.408670	2.446962	0.889759	0.167011	5.987624
Glavas 2003b	0.882389	0.304657	3.282380	2.896336	0.092816	10.774018
Stapleton 1995	0.706219	0.248734	4.020359	2.839254	0.061869	16.163287

References

Abramowitz M, Stegun IA. 1964. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. U.S. Department of Commerce, Washington, DC.

Begg CB, Berlin JA. 1988. Publication bias: a problem in interpreting medical data, *Journal of the Royal Statistical Society A* **151**: 419–445.

Bellavance F, Dionne G, Lebeau M. 2009. The value of a statistical life: a meta-analysis with a mixed effects regression model. *Journal of Health Economics* **28**: 444–464.

Cary B. 2008. Researchers find a bias towards upbeat findings on antidepressants. *New York Times*, January 17.

Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. 2004. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *Journal of the American Medical Association* **291**: 2457–65.

Cohen, J. 1988. Statistical Power Analysis in the Behavioral Sciences, 2nd edn. Erlbaum, Hillsdale.

Cooper HM, Hedges LV. (eds.) 1994. Handbook of Research Synthesis. Russell Sage, New York.

Copas J. 1999. What works? Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, A* **161**: 95–105.

Costa-Font J, Gammill M, Rubert G. 2011. Biases in the healthcare luxury good hypothesis: a meta-regression analysis, *Journal of the Royal Statistical Society A* **174**: 95–107.

Costa-Font J, McGuire A, Stanley TD. 2013. Publication selection in health policy research: the winner's curse hypothesis. *Health Policy* **109**: 78–87.

Davidson R, MacKinnon JG. 2004. Econometric Theory and Methods. Oxford University Press, Oxford.

Duval S, Tweedie RL. 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* **95**: 89–98.

Doucouliagos C(H), Stanley TD. 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations* **47**: 406–29.

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

Doucouliagos C(H), Stanley TD, Giles M. 2012 Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics* **31**: 197–206.

Egger M, Smith GD, Scheider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **316**: 629–34.

Feld LP, Heckemeyer JH. 2011. FDI and taxation: a meta-study. *Journal of Economic Surveys* **25**: 233–72.
Friedman RA. 2003. What you do know can't hurt you. New York Times, August 12.
Gordon RD. 1941. Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Annals of mathematical Statistics* **12**: 364–66.
Greene WE. 1990. Econometric Analysis. Macmillan, New York.
Havranek T. 2010. Rose effect and the Euro: is the magic gone?" *Review of World Economics* **146**: 241–261.
Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* **47**: 153–61.
Hedges LV, Olkin I. 1985. Statistical Methods for Meta-Analysis. Academic Press, Orlando.
Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* **21**: 1539–1558.
Hopewell S, Loudon K, Clarke M.J, Oxman A.D, Dickersin K. 2009. Publication bias in clinical trials due to statistical significance or direction of trial result. *Cochrane Review*, Issue 1. Available at http://www.thecochranelibrary.com.
Johnson N, Kotz S. 1970. Distributions in Statistics: Continuous Univariate Distribution. Wiley, New York.
Krakovsky M. 2004. Register or perish. *Scientific American* **291**: 18–20.
Laplace PS. 1812. Théorie Analytique des Probabilitiés. Gauthier Villars, Paris.
Lehrer J. 2010. The truth wears off: Is there something wrong with the scientific method? *The New Yorker*, December 13.
Mandel M, Rinott Y. 2009. A selection bias conflict and frequentist versus Bayesian viewpoints. *American Statistician* **63**: 211–217.
Moore DS. 1997. Bayes for beginners? Some reasons to hesitate. *The American Statistician* **51**: 254–261.
Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, Cooper NJ. 2009a. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology* **9**: 2. http://www.biomedcentral.com/1471-2288/9/2.
Moreno SG, Sutton AJ, Turner EH, Abrams KR, Cooper NJ, Palmer TM, Ades AE. 2009b. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. *BMJ* **339**:b2981: 494–98. DOI:10.1136/bmj.b2981.
Moreno SG, Sutton AJ, Thompson JR, Ades AE, Abrams KR, Cooper NJ. 2011. A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. *Statistics in Medicine* **31**: 1407–17.
Raudenbush SW. 1994. Random effects models, in The Handbook of Research Synthesis Cooper H, Hedges LV (eds). Russell Sage, New York, pp. 301–321.
Rosenthal R. 1979. The 'file drawer problem' and tolerance for null results. *Psychological Bulletin* **86**: 638–41.
Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. 2011. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* **12**: 122–42.
Senn S. 2008. A note concerning a selection 'paradox' of Dawid's. *American Statistician* **62**: 206–210.
Stanley TD. 2006. Two-stage precision-effect estimation and Heckman meta-regression for publication selection bias. School Working Paper, 2006–25, Deakin University, http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2006-25eco.pdf
Stanley TD. 2008. Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* **70**: 103–27.
Stanley TD, Doucouliagos C(H). 2007. Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. School Working Paper, Economics Series 2007–11, Deakin University. http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2007_11eco.pdf
Stanley TD, Jarrell SB, Doucouliagos H(C). 2010. Could it be better to discard 90% of the data? A statistical paradox. *American Statistician* **64**: 70–77.
Stanley TD, Doucouliagos H(C). 2012. Meta-Regression Analysis in Economics and Business. Routledge, Oxford, 2012.
Stanley TD, Doucouliagos C(H). 2013. Neither fixed nor random: weighted least squares meta-analysis. SWP, Economics Series 2013–1, Deakin University. http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2013_1.pdf
Stead LF, Perera R, Bullen C, Mant D, Lancaster T. 2008. Nicotine replacement therapy for smoking cessation, *The Cochrane Library*, Issue 2. Available at http://www.thecochranelibrary.com.
Sterling TD. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance. *Journal of the American Statistical Association* **54**: 30–34.
Sterling TD, Rosenbaum W.L, Weinkam JJ. 1995. Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician* **439**: 108–112.
Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* **358**: 252–60.
Viscusi WK. 1993. The value of risks to life and health. *Journal of Economic Literature* **31**: 1912–46.
Wooldridge JM. 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.