Research Article

Statistics in Medicine

Received 3 March 2013, Accepted 27 February 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6481

Neither fixed nor random: weighted least squares meta-analysis

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

This study challenges two core conventional meta-analysis methods: fixed effect and random effects. We show how and explain why an unrestricted weighted least squares estimator is superior to conventional random-effects meta-analysis when there is publication (or small-sample) bias and better than a fixed-effect weighted average if there is heterogeneity. Statistical theory and simulations of effect sizes, log odds ratios and regression coefficients demonstrate that this unrestricted weighted least squares estimator provides satisfactory estimates and confidence intervals that are comparable to random effects when there is no publication (or small-sample) bias and identical to fixed-effect meta-analysis when there is no heterogeneity. When there is publication selection bias, the unrestricted weighted least squares approach dominates random effects; when there is excess heterogeneity, it is clearly superior to fixed-effect meta-analysis. In practical applications, an unrestricted weighted least squares weighted average will often provide superior estimates to both conventional fixed and random effects.

Keywords: meta-analysis; meta-regression; weighted least squares; fixed effect; random effects

1. Introduction

Nearly all meta-analyses report a 'fixed-effect' or a 'random-effects' weighted average, often both [1,2]. However, it is widely known that the fixed-effect estimator produces confidence intervals with poor coverage when applied to unconditional inference, that is, to populations that may not be entirely identical to the one sampled [2–4]. Random effects, on the other hand, are highly sensitive to the accuracy of the estimate of the between-study variance, τ^2 [1], and conventional estimates of τ^2 are biased [3]. When there is publication (or small-sample) bias, random effects have larger biases than fixed effect [4–8].

In this paper, we propose the routine use of a simple unrestricted weighted least squares meta-regression that offers the best of both approaches. We show how this unrestricted weighted least squares estimator corrects the poor coverage of the fixed-effect estimator. Further, when there is either publication selection or small-sample bias, our simulations demonstrate that the unrestricted weighted least squares dominate random-effects meta-analysis, whether the reviewer is synthesizing RCTs or regression estimates.

In practice, our approach addresses the same problems of conventional meta-analysis as does Henmi and Copas [4]. Their hybrid confidence interval is centered on the fixed-effect estimate, same as in our weighted least squares estimator, but Henmi and Copas calculate its width from the random-effects setting, further taking into account the uncertainty of estimating τ^2 [4]. We believe that our unrestricted weighted least squares approach is more simple and elegant. Weighted least squares have a long history with well-established statistical properties rooted in the Gauss–Markov Theorem [9–12]. Weighted least squares confidence intervals are easily and automatically calculated by regression routines found in all standard statistical software.

Weighted least squares have been used by many meta-analysts in different contexts [1, 13–22]. We fully recognize that weighted least squares are an integral component of all of these methods, including

[a]*Department of Economics, Hendrix College, 1600 Washington St., Conway, AR, 72032, U.S.A.*
[b]*Department of Economics, Deakin University, 221 Burwood Highway, Burwood, 3125, Victoria, Australia*
[*]*Correspondence to: T. D. Stanley, Julia Mobley Professor of Economics, Hendrix College, 1600 Washington St., Conway, AR, 72032, U.S.A.*
[†]*E-mail: Stanley@hendrix.edu*

fixed and random effects. However, the key difference among these methods lies in exactly how each implements weighted least squares, and these differences matter.

To our knowledge, no one has suggested that an unrestricted weighted least squares (WLS) should replace random-effects meta-analysis. Nor has anyone demonstrated the superiority of an unrestricted weighted least squares over conventional meta-analysis. However, particle physicists have long used a similar weighted least squares approach for experimental measurements of the mass and charge of fundamental particles (*e.g.*, bosons, leptons, quarks) without ever referring to meta-analysis [23]. Rather than embracing these methods, meta-analysts have thus far denied their relevance. 'Our model-based analysis shows that the conventional additive random-effects model appears to fit the data better than the multiplicative model, so our suggestion is that here the (particle physicists) might consider changing their practice' [23, p. 120]. We demonstrate just the opposite through realistic simulations of meta-analyses across a wide array of designs, conditions, and effect measures. Our simulations show that weighted least squares estimates are often superior to random effects even when we are confined to the conventional additive random-effects model. Thus, meta-analysts would do well to report this unrestricted weighted least squares estimate routinely as a summary statistic.

2. Simple weighted averages

The fixed-effect estimator assumes that the individual reported effects, y_i, are a random draw from a normal population. Or

$$y_i = \mu + \varepsilon_i \quad \text{and} \quad \varepsilon_i \sim \mathrm{N}\left(0, \sigma_i^2\right) \quad \text{for i} = 1, 2, \ldots, m; \tag{1}$$

where m is the number of reported estimates to be meta-analyzed. Random effects allow individual means to vary randomly around μ, the population parameter of interest or the treatment effect.

$$y_i = \mu + \theta_i + \varepsilon_i \quad ; \quad \theta_i \sim \mathrm{N}\left(0, \tau^2\right) \text{ and } \varepsilon_i \sim \mathrm{N}\left(0, \sigma_i^2\right) \quad \text{for i} = 1, 2, \ldots, m. \tag{2}$$

All three estimators: fixed effect, random effects, and unrestricted weighted least squares model, reported effects as

$$y_i \sim N\left(\mu, v_i\right) \tag{3}$$

with different assumptions about the individual variances, v_i. Random effects assume that variances are additive: $v_i = \left(\sigma_i^2 + \tau^2\right)$, where τ^2 is the usual between-study or heterogeneity variance. Fixed effect assumes that there is no excess heterogeneity, or $\tau^2 = 0$. The unrestricted weighted least squares can also be modeled by equation (1); however, it assumes only that the variances can be estimated up to some unknown multiplicative constant ϕ, or that $v_i = \phi\sigma_i^2$.

The Gauss-Markov theorem proves that as long as v_i is known up to some proportional constant, ϕ, the conventional weighted least squares estimator provides the best (minimum variance) linear unbiased estimator [10, 11]. With consistent estimates of σ_i^2 (such as each study's squared standard error), weighted least squares provide consistent, asymptotically efficient and asymptotically normal estimates [12].

All three estimators can be written in a common compact form:

$$\hat{\mu} = \quad \Sigma w_i y_i / \Sigma w_i \tag{4}$$

However, each employs different weights, w_i, and thereby has different variances. Fixed effect uses weights, $w_i = 1/\sigma_i^2$, and has variance, $1/\Sigma w_i$. Random effects has weights, $w_i' = 1/\left(\sigma_i^2 + \tau^2\right)$ with variance, $1/\Sigma w_i'$. Lastly, the unrestricted weighted least squares' weights are $w_i^* = 1/\left(\phi\sigma_i^2\right)$ with variance $= 1/\Sigma w_i^* = \phi/\Sigma 1/\sigma_i^2$.

Thus, fixed effect, $\hat{\mu}_F$, and the unrestricted weighted least squares estimators, $\hat{\mu}_W$, are identical. Substituting $1/\sigma_i^2$ for w_i into equation (4) implies that

$$\hat{\mu}_F = \Sigma\left(1/\sigma_i^2\right) y_i / \Sigma\left(1/\sigma_i^2\right) = (1/\phi)\Sigma\left(1/\sigma_i^2\right) y_i / (1/\phi)\Sigma\left(1/\sigma_i^2\right) = \hat{\mu}_W; \text{ for all } \phi \neq 0. \tag{5}$$

However, $\hat{\mu}_F$ and $\hat{\mu}_W$ have different variances. The variance of $\hat{\mu}_W$ is ϕ times the variance of $\hat{\mu}_F$.

Sample estimates are easy to obtain for all of the previously mentioned parameters from the conventional information collected in a systematic review. First, the standard error of each study's reported

estimate, SE_i, may be used in place of σ_i. Second, ϕ is automatically estimated from the meta-sample by conventional weighted least squares statistical software (e.g., STATA). Ordinary least squares will also correctly calculate $\hat{\mu}_W$ and its confidence interval. To do so, run a simple meta-regression of the standardized effect size, $t_i = y_i/SE_i$, with precision, $1/SE_i$, as the independent variable and no intercept [14]. The mean squared error of this simple regression,

$$H^2 = \sum [t_i - (\hat{\mu}_W/SE_i)]^2/(m-1), \tag{6}$$

serves as an estimate of ϕ and is automatically employed to help calculate the standard error and confidence interval of $\hat{\mu}_W$. Both H and $I^2 = (H^2 - 1)/H^2$ are used to measure heterogeneity in systematic reviews [22]. Lastly, τ^2 is routinely calculated by a separate algorithm, often the method of moments or the DerSimonian-Laird method [3, 24]. Unlike the random-effects' estimate of τ^2, unrestricted weighted least squares' estimate of ϕ is not restricted in any way. It can, therefore, be less than one if a research literature has smaller heterogeneity than what is consistent with the reported standard errors, while $\hat{\tau}^2$ is usually truncated at 0. Because ϕ is estimated, $\hat{\mu}_W$ will be distributed as a Student's t-test with m-1 degrees of freedom.

In the next section, we offer realistic simulations of these three estimators. Simulations are needed because variances must be estimated in applications. As a result, the statistical properties of random effects are known only asymptotically, and random effects are very sensitive to the biased estimate of the heterogeneity variance [1, 3]. Also, our simulations do not follow the multiplicative variance structure needed by the Gauss–Markov theorem to guarantee weighted least squares' optimal properties. In these simulations, excess heterogeneity is always introduced as an additive term, that is, as assumed by the random-effects model (2). We take it for granted that the additive model, equation (2), is more realistic in actual applications. The purpose of this paper is not to show that the weighted least squares model is somehow better than the random-effects model. Rather, we demonstrate in the next pages that this weighted least squares approach is robust to violations to its assumptions and still has comparable or superior statistical properties over a wide range conditions, designs, and effect size measures, including log odds ratios and regression coefficients.

3. Simulations

We simulate a wide variety of conditions across different measures of effect sizes and designs: standardized mean differences, log odds ratios (both from randomized controlled trials), and regression coefficients from observational studies. Our simulations begin with standardized mean differences from randomized controlled trials calibrated around a well-known meta-analysis of the FDA's registry of the clinical trials of antidepressants [25]. We commence with standardized mean differences because they have well-behaved distributions, are used widely across the disciplines, and can be easily transformed into other measures of effect size. To ensure the generalizability of our findings, log odds ratios from randomized controlled trials are also simulated in Section 3.2.

3.1. Standardized mean differences from randomized controlled trials

The average Cohen's d for published antidepressants trials is 0.47, which we round to 0.5 in our simulations [6, p.74]. Recall that Cohen's d is the standardized difference in sample means with the pooled standard deviation in the denominator. However, there is clear evidence that published antidepressant trials have substantial publication selection, small-sample, or reporting bias [6, 25]. Thus, the true effect is likely to be much smaller. To encompass this and most other meta-analyses, we also simulate a small effect, $d = 0.2$ and no effect, $d = 0$, in addition to this medium effect size, $d = 0.5$. We simulate both the case of 50% publication selection to be consistent with what we observe among antidepressant trials and also the case of no publication bias to capture other medical meta-analyses as well. We choose meta-analysis sample sizes, $m = \{5, 10, 20, 40, 80\}$, both larger and smaller than Turner, *et al.*, [25], because smaller sample sizes are common in the meta-analysis of medical research.

To be more precise, the simulations first involved the generation of outcomes as

$$y_{cj} = x_{cj} + u_j \quad \text{j} = 1, 2, \ldots, n \tag{7}$$

for the control group; where $u_j \sim N(0, 50^2)$ and $x_{cj} \sim N(300, 86.6^2)$. Outcomes in the experimental group are generated in the exact same way except that there is an added treatment effect:

$$T_e = \mu + \theta_i \tag{8}$$

Where $\theta_i \sim \mathrm{N}\left(0, \sigma_h^2\right)$. In the simulations we assume there is either no effect, a small effect, $\mu = 20$, or a medium size effect, $\mu = 50$; thus, Cohen's d is $\{0, 0.2, 0.5\}$. The sample sizes for each group are set at: 32, 64, 125, 250, or 500 subjects. The simulated effects then become an input into simulated meta-analyses with sample sizes m $= \{5, 10, 20, 40, 80\}$. We investigate a full range of heterogeneity by varying the standard deviation of the random heterogeneity, $\sigma_h = \{0, 6.25, 12.5, 25, 50, 100, 200\}$, because heterogeneity is quite common in medical research and to encompass all meta-analyses across the different disciplines [26]. See Tables I and II.

Table I. Coverage of experimental results, standardized mean differences, ($d = 0$).

m: Meta-analysis sample size	Random heterogeneity (σ_h)*	$I^{2\dagger}$	FEMA	REMA	WLS
5	0	0.1293	0.9509	0.9627	0.9455
5	6.25	0.2146	0.8725	0.9187	0.9322
5	12.5	0.3922	0.6997	0.8612	0.9090
5	25	0.6798	0.4309	0.8131	0.8807
5	50	0.8863	0.2323	0.8012	0.8794
5	100	0.9661	0.1294	0.7867	0.8794
5	200	0.9886	0.0739	0.7541	0.8960
10	0	0.1099	0.9540	0.9650	0.9509
10	6.25	0.2245	0.8720	0.9265	0.9284
10	12.5	0.4737	0.6938	0.8926	0.8971
10	25	0.7808	0.4391	0.8885	0.8764
10	50	0.9341	0.2441	0.8831	0.8706
10	100	0.9803	0.1376	0.8570	0.8741
10	200	0.9931	0.0937	0.8265	0.8970
20	0	0.0878	0.9488	0.9595	0.9489
20	6.25	0.2284	0.8721	0.9305	0.9269
20	12.5	0.5324	0.6931	0.9212	0.8972
20	25	0.8232	0.4370	0.9207	0.8719
20	50	0.9472	0.2455	0.9116	0.8645
20	100	0.9841	0.1443	0.8960	0.8828
20	200	0.9941	0.0948	0.8609	0.9027
40	0	0.0664	0.9504	0.9601	0.9484
40	6.25	0.2422	0.8728	0.9367	0.9232
40	12.5	0.5646	0.6963	0.9362	0.8937
40	25	0.8406	0.4413	0.9374	0.8723
40	50	0.9526	0.2472	0.9265	0.8675
40	100	0.9855	0.1401	0.9121	0.8832
40	200	0.9944	0.0963	0.8753	0.9104
80	0	0.0485	0.9492	0.9562	0.9480
80	6.25	0.2479	0.8735	0.9422	0.9261
80	12.5	0.5824	0.6885	0.9416	0.8887
80	25	0.8480	0.4475	0.9448	0.8727
80	50	0.9547	0.2505	0.9364	0.8712
80	100	0.9860	0.1501	0.9229	0.8835
80	200	0.9945	0.0998	0.8807	0.9167
Average			0.4904	0.9013	0.9005
Mean absolute deviation from 0.95			0.4599	0.0517	0.0496

*σ_h is the standard deviation of the random heterogeneity.
†I^2 is the proportion of the total variation among the empirical effects that is attributable to heterogeneity [22]. FEMA, REMA and WLS denote the fixed-effect, random-effects and unrestricted weighted least squares meta-analysis averages, respectively.

Table II. Coverage of experimental results, standardized mean differences, ($d = 0.5$).

m: Meta-analysis sample size	Random heterogeneity (σ_h)*	$I^{2\dagger}$	FEMA	REMA	WLS
5	0	0.1316	0.9508	0.9640	0.9485
5	6.25	0.2096	0.8710	0.9174	0.9314
5	12.5	0.3861	0.7012	0.8608	0.9058
5	25	0.6796	0.4362	0.8248	0.8914
5	50	0.8830	0.2325	0.7987	0.8726
5	100	0.9659	0.1387	0.7862	0.8794
5	200	0.9887	0.0821	0.7540	0.8949
10	0	0.1087	0.9512	0.9643	0.9498
10	6.25	0.2200	0.8749	0.9237	0.9290
10	12.5	0.4645	0.7011	0.8936	0.8963
10	25	0.7742	0.4438	0.8825	0.8723
10	50	0.9314	0.2478	0.8769	0.8635
10	100	0.9800	0.1377	0.8613	0.8661
10	200	0.9929	0.0926	0.8235	0.8810
20	0	0.0853	0.9480	0.9573	0.9452
20	6.25	0.2247	0.8745	0.9323	0.9254
20	12.5	0.5237	0.7024	0.9198	0.8957
20	25	0.8191	0.4448	0.9201	0.8699
20	50	0.9461	0.2547	0.9176	0.8643
20	100	0.9838	0.1373	0.8977	0.8661
20	200	0.9940	0.0940	0.8624	0.8779
40	0	0.0648	0.9509	0.9583	0.9485
40	6.25	0.2363	0.8739	0.9360	0.9232
40	12.5	0.5564	0.6997	0.9331	0.8912
40	25	0.8368	0.4329	0.9335	0.8626
40	50	0.9508	0.2438	0.9318	0.8528
40	100	0.9851	0.1372	0.9184	0.8554
40	200	0.9944	0.0835	0.8731	0.8546
80	0	0.0487	0.9493	0.9559	0.9485
80	6.25	0.2420	0.8689	0.9392	0.9220
80	12.5	0.5744	0.7014	0.9407	0.8933
80	25	0.8436	0.4433	0.9427	0.8666
80	50	0.9537	0.2358	0.9406	0.8450
80	100	0.9858	0.1197	0.9179	0.8097
80	200	0.9945	0.0698	0.8796	0.7984
Average			0.4894	0.9011	0.8885
Mean absolute deviation from 0.95			0.4608	0.0517	0.0615

*σ_h is the standard deviation of the random heterogeneity.
†I^2 is the proportion of the total variation among the empirical effects that is attributable to heterogeneity [22]. FEMA, REMA and WLS denote the fixed-effect, random-effects and unrestricted weighted least squares meta-analysis averages, respectively.

Tables I and II display the coverage rates for the standardized mean differences as measured by Cohen's d. Hedges' g, which applies a gamma correction to Cohen's d [27], was also simulated, but the differences were inconsequential. Tables I and II also report the percentage of unexplained random heterogeneity found among the estimated effects, relative to the total variation in observed effects, I^2 [22]. The reported values of I^2 are calculated 'empirically' for each replication of these simulations and averaged. When $\sigma_h^2 = 0$, the 'true' I^2 would also be zero; however, the conventional truncation of I^2 at zero imparts an upward bias. 95% confidence intervals are constructed for each replication using the formulas and methods reported in Section 2, above. Lastly, the coverage proportions from 10,000 replications are reported in the last three columns of Table I and II for d = {0 and 0.5, respectively} . Results for $d = 0.2$ may be found in an internet appendix at: http://www.deakin.edu.au/business/economics/research/meta-analysis.

As expected, the coverage probabilities are very poor for the conventional fixed-effect meta-analysis (FEMA) when there is excess heterogeneity. One might excuse fixed effect in these cases, because it is not

designed for unconditional inference; that is, for populations that differ in any way from the one sampled [3]. However even when FEMA's model is true (i.e., $\sigma_h^2 = 0$), the unrestricted weighted least squares (WLS) provide practically equivalent coverage. Across Tables I, II and Appendix Table I, both estimators depart from the .95 nominal level by .002, on average, when $\sigma_h^2 = 0$. However, the central finding of these simulations is that the unrestricted weighted least squares variances make an acceptable allowance for the actual uncertainty of the fixed-effect estimate, regardless of the level of excess heterogeneity. Tables I, II and Appendix I show that WLS's coverage rates are slightly closer to the nominal level of 95% than random effects meta-analysis (REMA), while the reverse is true when d = 0.5 (Table II). On average and across Tables I, II and Appendix Table I, weighted least squares' coverage rates depart from the nominal level of 95% by 5.44%; whereas random effects are off by 5.19%. From a practical point of view, these coverage rates are equivalent.

We do not report bias and mean square errors (MSE) for these alternative meta-analysis estimators because the properties of the conventional meta-analysis estimators are already well known. Weighted least squares' point estimate is identical to the conventional fixed effect meta-analysis; thus, they must have the same bias and MSE. Weighted least squares differ from the conventional fixed effect only in their variances. With heterogeneity, weighted least squares will thereby correctly have wider confidence intervals and larger p-values than fixed effect.

In previous studies, fixed effect has been shown to be less biased than random effects when there is publication selection for statistical significance (or small-sample bias) [4–8]. Thus, the real advantage of weighted least squares approach over random effects is seen when there is publication selection (or small-sample) bias. Tables III, IV and Appendix Table II report the bias and MSE for weighted least

Table III. Bias and MSE of experimental results with 50% publication bias, standardized mean differences, (d = 0).

m: Meta-analysis sample size	Random heterogeneity (σ_h^*)	REMA Bias	WLS Bias	REMA MSE	WLS MSE
10	0	0.1274	0.1064	0.0168	0.0119
10	6.25	0.1366	0.1139	0.0195	0.0140
10	12.5	0.1556	0.1293	0.0258	0.0189
10	25	0.2024	0.1695	0.0457	0.0362
10	50	0.2996	0.2597	0.1069	0.0940
10	100	0.4908	0.4230	0.3057	0.2662
10	200	0.8922	0.7009	1.0567	0.7362
20	0	0.1277	0.1065	0.0166	0.0116
20	6.25	0.1364	0.1133	0.0190	0.0133
20	12.5	0.1558	0.1284	0.0251	0.0176
20	25	0.2030	0.1693	0.0436	0.0324
20	50	0.3001	0.2561	0.0985	0.0787
20	100	0.4927	0.4221	0.2755	0.2208
20	200	0.8885	0.6937	0.9210	0.5941
40	0	0.1275	0.1063	0.0164	0.0114
40	6.25	0.1366	0.1132	0.0189	0.0131
40	12.5	0.1567	0.1290	0.0250	0.0172
40	25	0.2044	0.1695	0.0430	0.0306
40	50	0.3006	0.2572	0.0947	0.0728
40	100	0.4933	0.4215	0.2602	0.1990
40	200	0.8909	0.6831	0.8608	0.5229
80	0	0.1281	0.1066	0.0165	0.0114
80	6.25	0.1370	0.1134	0.0189	0.0130
80	12.5	0.1569	0.1289	0.0248	0.0169
80	25	0.2042	0.1699	0.0423	0.0298
80	50	0.2996	0.2561	0.0919	0.0688
80	100	0.4949	0.4229	0.2532	0.1893
80	200	0.8886	0.6808	0.8221	0.4907
Average bias or MSE		0.3296	0.2697	0.1987	0.1369

MSE, mean square error; REMA, random-effects meta-analysis; WLS, weighted least square.
*σ_h is the standard deviation of the random heterogeneity.

Table IV. Bias and MSE of experimental results with 50% publication bias, standardized mean differences, ($d = 0.5$).

m: Meta-analysis sample size	Random heterogeneity (σ_h^*)	REMA Bias	WLS Bias	REMA MSE	WLS MSE
10	0	0.0069	0.0054	0.0011	0.0011
10	6.25	0.0089	0.0057	0.0017	0.0017
10	12.5	0.0161	0.0076	0.0034	0.0038
10	25	0.0468	0.0260	0.0092	0.0110
10	50	0.1396	0.0936	0.0398	0.0415
10	100	0.3344	0.2275	0.1822	0.1524
10	200	0.7339	0.4419	0.8122	0.4711
20	0	0.0065	0.0055	0.0006	0.0005
20	6.25	0.0086	0.0057	0.0009	0.0009
20	12.5	0.0169	0.0079	0.0018	0.0019
20	25	0.0496	0.0258	0.0059	0.0057
20	50	0.1390	0.0901	0.0293	0.0242
20	100	0.3345	0.2125	0.1474	0.0951
20	200	0.7286	0.4087	0.6698	0.3028
40	0	0.0059	0.0053	0.0003	0.0003
40	6.25	0.0082	0.0055	0.0005	0.0005
40	12.5	0.0168	0.0071	0.0010	0.0010
40	25	0.0493	0.0249	0.0041	0.0031
40	50	0.1386	0.0877	0.0244	0.0157
40	100	0.3342	0.2125	0.1298	0.0695
40	200	0.7281	0.4021	0.5995	0.2252
80	0	0.0057	0.0053	0.0002	0.0002
80	6.25	0.0082	0.0056	0.0003	0.0002
80	12.5	0.0174	0.0076	0.0007	0.0005
80	25	0.0505	0.0258	0.0034	0.0019
80	50	0.1393	0.0879	0.0219	0.0117
80	100	0.3316	0.2085	0.1191	0.0555
80	200	0.7280	0.3988	0.5647	0.1906
Average bias or MSE		0.1833	0.1089	0.1205	0.0603

MSE, mean square error; REMA, random-effects meta-analysis; WLS, weighted least square.
*σ_h is the standard deviation of the random heterogeneity.

squares and random effects when half of the studies selectively report significantly positive results. For the other half, the first estimate that is randomly generated is reported, and the simulation design is identical to what is described above and used to generate Tables I, II and Appendix Table I. For the selected 50%, everything is randomly generated as before, except all of the random generating processes are repeated over and over again, until an estimated effect is statistically positive. In *all* cases, weighted least squares have smaller bias than random effects when there is publication selection (or small-sample) bias—see Tables III, IV and Appendix Table II. In some cases, the difference is of no practical import; in others, random effects have a notably larger, practically significant bias, as much as 0.3 standard deviation and more.

On average, our simulations find that random-effects' bias is 68% larger than weighted least squares meta-analysis for $d = 0.5$ (Table IV), 33% larger for $d = 0.2$, and 22% for $d = 0$. Likewise random effects' MSE is typically much larger than unrestricted weighted least squares' MSE. In only 5% of these cases does REMA have smaller MSE than WLS, and the differences are practically negligible for those cases. On the other hand, REMA has twice the MSE as weighted least squares for $d = 0.5$, 64% larger for $d = 0.2$, and 45% for $d = 0$, on average. When there is publication selection bias, the unrestricted weighted least squares estimator dominates random effects.

3.2. Log odds ratios

We also simulate binary RCTs as measured by log odds ratios, because they are common in medical research. This second set of simulations are calibrated around another actual medical meta-analysis, Stead *et al*. [28]. Among 112 RCTs of nicotine replacement therapy (NRT), the average increase in the quit

rate is 6% from a base of approximately 10% in the control group [28]. Because there is clear evidence of publication bias (or small-sample bias) and the true effect is likely to be much smaller than what is reported [6], we simulate smaller effect sizes than the 6% average: a 3% increase in smoking cessation for the treated group and a 0% quit rate. We calibrate I^2 in our simulations for this middle effect (3% increase in quitting) to match exactly the I^2 value found among NRT trials (30%). Because standard errors of log odds ratios depend on the magnitude of the probabilities (or rates of smoking cessation), heterogeneity in the true probability of smoking cessation will cause I^2 to be larger (smaller) for smaller (larger) effects.

For the experimental group, our simulations set the probability of a successful outcome (smoking cessation) to be

$$P = 0.10 + effect + u; \tag{9}$$

where effect $= \{0, 0.03, 0.06\}$ and heterogeneity is induced by $u \sim N(0, .006^2)$. The variance of u is chosen to make the I^2 exactly 30% when the true effect is 0.03 and there is publication bias—see Table VI. The probability that a member of the control group has a successful outcome is held constant at 0.10. Binary outcomes are generated with the above probabilities randomly using either 50, 100, 250, or 500 subjects. These sample sizes and their relative frequencies ($n = 100$ is doubly weighted) were selected to reflect the actual sample sizes reported among NRT trials. After calculating a log odds ratio from these randomly generated groups, above, it becomes an input into each of 10,000 simulated meta-analyses with sample sizes, $m = \{5, 10, 20, 40, 80\}$.

Table V reports the confidence intervals for these alternative estimators when there is no publication bias. As before, the unrestricted weighted least squares has better coverage than fixed effect, and the difference between random effects and weighted least squares is practically negligible. On average, REMA is 0.0003 closer to the nominal level. When there is no publication selection bias, binary outcomes, and realistic parameters from an actual meta-analysis, the coverage of the unrestricted weighted least squares is practically as good as or better than conventional meta-analysis methods.

When there is publication bias (Table VI), the unrestricted weighted least squares dominates random effects. In *all* cases, the bias and the MSE of the unrestricted weighted least squares are smaller, on average, than for random effects (see Table VI). One might point out that the improvement of the unrestricted weighted least squares here is modest. REMA's MSE is only 39% larger and its bias is 18% greater. However, when there are high levels of heterogeneity (recall Tables III, IV, and Appendix Table II), the improvement of weighted least squares over random effects can be much larger. Random estimation errors cause REMA to have a smaller bias or a smaller MSE only 2.5% of the time, if we look at each of these 120,000 meta-analyses simulated in Table VI, separately.

Table V. Coverage for log odds ratios; no publication bias.

m: Meta-analysis Sample Size	True log odds ratio	$I^{2\dagger}$	FEMA	REMA	WLS
5	0	0.2805	0.8117	0.8943	0.9262
5	0.30	0.2407	0.8460	0.9092	0.9337
5	0.54	0.2199	0.8698	0.9224	0.9386
10	0	0.3225	0.8259	0.9156	0.9264
10	0.30	0.2680	0.8506	0.9222	0.9294
10	0.54	0.2307	0.8703	0.9259	0.9261
20	0	0.3542	0.8219	0.9255	0.9223
20	0.30	0.2857	0.8465	0.9249	0.9232
20	0.54	0.2394	0.8692	0.9321	0.9284
40	0	0.3801	0.8137	0.9305	0.9163
40	0.30	0.3028	0.8470	0.9343	0.9222
40	0.54	0.2490	0.8615	0.9315	0.9207
80	0	0.3981	0.8227	0.9437	0.9201
80	0.30	0.3222	0.8440	0.9386	0.9269
80	0.54	0.2624	0.8610	0.9380	0.9229
	Average		0.8441	0.9259	0.9256

$^{\dagger}I^2$ is the proportion of the total variation among the empirical effects that is attributable to heterogeneity [22]. FEMA, REMA and WLS denote the fixed-effect, random-effects and unrestricted weighted least squares meta-analysis averages, respectively.

Table VI. Bias and MSE for log odds ratios with 50% publication selection.

m: Meta-analysis sample size	True log odds ratio	$I^{2\dagger}$	REMA Bias	WLS Bias	REMA MSE	WLS MSE
10	0	0.5186	0.4161	0.3476	0.1791	0.1276
10	0.30	0.2977	0.2447	0.2090	0.0659	0.0505
10	0.54	0.1668	0.1311	0.1126	0.0242	0.0200
20	0	0.5195	0.4163	0.3471	0.1762	0.1238
20	0.30	0.2974	0.2431	0.2073	0.0619	0.0464
20	0.54	0.1512	0.1283	0.1114	0.0199	0.0161
40	0	0.5193	0.4165	0.3473	0.1750	0.1223
40	0.30	0.2997	0.2433	0.2076	0.0606	0.0448
40	0.54	0.1354	0.1266	0.1118	0.0177	0.0143
80	0	0.5195	0.4167	0.3469	0.1744	0.1212
80	0.30	0.3049	0.2430	0.2064	0.0598	0.0434
80	0.54	0.1284	0.1264	0.1125	0.0168	0.0135
Average bias or MSE		0.3215	0.2627	0.2223	0.0860	0.0620

$^{\dagger}I^2$ is the proportion of the total variation among the empirical effects that is attributable to heterogeneity [22]. REMA and WLS denote the random-effects and unrestricted weighted least squares meta-analysis averages, respectively.

3.3. Regression estimates

Our final set of simulations concern estimated regression coefficients from observational studies. Regression is commonly used in many disciplines, including health and epidemiology, and it encompasses many other statistical tests, including ANOVA, t-tests, and quasi-experimental designs (regression discontinuity, instrumental variables, difference-in-difference) [5, 29]. The basic simulation design here follows a series of prior studies [5, 6, 16]. Greater detail can be found in Stanley and Doucouliagos [30]. First data are randomly generated and used to estimate the target regression coefficient, α_1, from

$$Z_j = 100 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + u_j \qquad j = 1, 2, \ldots, n \tag{10}$$

Where $u_j \sim N(0, 100^2)$, $X_{1j} \sim$ Uniform (100,300), and α_1 =1. When $\alpha_1 = 1$, the correlation between Z and X_1 is 0.27, a small effect size by conventional guidelines [31]. A wide range of sample sizes are assumed to be used to estimate α_1 in the primary literature, $n = \{62, 125, 250, 500, 1000\}$, and as before, we allow for broad range in meta-analysis sample sizes $\{5, 10, 20, 40, 80\}$.

X_{2j} is equal to X_{1j} plus an $N(0,50^2)$ disturbance. When a relevant variable, like X_2, is omitted from a regression but is correlated with the included independent variable, like X_1, the estimated regression coefficient ($\hat{\alpha}_{1i}$) will be biased. This omitted-variable bias is $\alpha_2 \cdot \alpha_{12}$; where $\alpha_{12} = 1$ is the slope coefficient of a regression of X_{2j} on X_{1j}. In these simulations, α_2 is generated randomly for each study, $\alpha_{2i} \sim N\left(0, \sigma_h^2\right)$. That is, empirical effects are assigned random additive heterogeneity just as assumed by random effects with variance = σ_h^2. Values of random heterogeneity, σ_h, were selected to encompass the heterogeneity found in past meta-analyses. For example, among minimum-wage elasticities, I^2 is 90% [32]; it is 93% among estimates of the value of statistical life [33] and 97% among the partial correlations of CEO pay and corporate performance [34].

The coverage proportions for these alternative meta-analysis estimators over 10,000 replications are reported in the last three columns of Appendix Table III. The coverage proportions vary by 0.003, or less, from one simulation of 10,000 to the next simulation of 10,000 replications.

As before, the coverage probabilities are very poor for the conventional fixed-effect meta-analysis when there is excess heterogeneity; however, the coverage of the unrestricted weighted least squares is better than random-effects' coverage. In over two-thirds of the cases, WLS's coverage is closer to the nominal 0.95 level. On average, it is within 3.66% of the nominal level, while random effects are off by twice as much, 7.36%. The unrestricted weighted least squares' coverage is comparable or practically superior to random effects when there is no publication bias.

Table VII reports the bias and MSE when half of the studies selectively report significantly positive results. In all cases, weighted least squares have less bias than random effects when there is publication selection (or small-sample) bias. On average, our simulations find that random-effects' bias is 77% larger

Table VII. Bias and MSE of regression estimates with 50% publication selection ($\alpha_1 = 1$).

m: Meta-analysis sample size	Random heterogeneity (σ_h)*	REMA Bias	WLS Bias	REMA MSE	WLS MSE
10	0	0.0086	0.0066	0.0032	0.0031
10	0.125	0.0131	0.0085	0.0057	0.0059
10	0.25	0.0255	0.0124	0.0115	0.0138
10	0.50	0.0823	0.0498	0.0324	0.0379
10	1.0	0.2546	0.1783	0.1415	0.1257
10	2.0	0.6186	0.4143	0.6400	0.3879
10	4.0	1.3651	0.7448	2.8433	1.0700
20	0	0.0080	0.0067	0.0016	0.0016
20	0.125	0.0118	0.0075	0.0030	0.0030
20	0.25	0.0251	0.0120	0.0059	0.0068
20	0.50	0.0858	0.0522	0.0200	0.0206
20	1.0	0.2578	0.1756	0.1050	0.0761
20	2.0	0.6206	0.3940	0.5160	0.2561
20	4.0	1.3615	0.7048	2.3395	0.7234
40	0	0.0071	0.0063	0.0008	0.0008
40	0.125	0.0114	0.0074	0.0015	0.0015
40	0.25	0.0263	0.0126	0.0033	0.0035
40	0.50	0.0852	0.0501	0.0135	0.0113
40	1.0	0.2571	0.1733	0.0854	0.0519
40	2.0	0.6209	0.3856	0.4523	0.1983
40	4.0	1.3629	0.6745	2.1020	0.5596
80	0	0.0072	0.0067	0.0004	0.0004
80	0.125	0.0119	0.0079	0.0008	0.0008
80	0.25	0.0264	0.0124	0.0020	0.0018
80	0.50	0.0861	0.0503	0.0105	0.0070
80	1.0	0.2582	0.1746	0.0761	0.0415
80	2.0	0.6237	0.3809	0.4221	0.1689
80	4.0	1.3692	0.6653	1.9960	0.4936
Average bias or MSE		0.3390	0.1920	0.4227	0.1526

*σ_h is the standard deviation of the random heterogeneity.
MSE, mean square error; REMA, random-effects meta-analysis; WLS, weighted least square.

than weighted least squares, and random-effects' MSE is just under three times larger than weighted least squares' MSE. As other studies and simulations have shown, when there is publication selection or small-sample bias, random effects give an unacceptable summary of research findings

4. Conclusions

Publication bias is common in medical research [35]. Unfortunately, tests for publication and small-sample bias are well known to have low power [16,36]. Thus, it is prudent for meta-analysts to assume that there is publication (or small-sample) bias, regardless of what their tests might indicate. With unrestricted weighted least squares so clearly dominating random effects when there is publication (or small-sample) bias and given that the unrestricted weighted least squares' confidence intervals are practically equivalent or superior to random effects when there is no publication bias, we see little reason why unrestricted weighted least squares should not be reported in all meta-analyses.

Weighted least squares are well grounded by statistical theory, the Gauss–Markov theorem, and are very easy to implement. One need merely to run a simple ordinary least squares regression of the estimate's standardized value y_i/SE_i, against its precision ($1/SE_i$) with no intercept. All regression software will correctly calculate this simple substitute for random effects, its standard error, its t-test, and its confidence interval. Nothing further is required.

The contribution of this paper is quite modest but potentially far reaching. At a minimum, we offer a simple correction for the standard errors of conventional fixed-effect meta-analysis when applied to unconditional inferences and demonstrate how the resulting unrestricted weighted least squares estimator is practically comparable and often superior to conventional random-effects meta-analysis. In practice,

the difference between these meta-analysis estimates can be very large with important practical consequences. For example, among the estimates of the value of a statistical life, random-effects meta-analysis estimates this key policy parameter to be more than three times larger ($5.7 million) than the unrestricted weighted least squares estimate ($1.8 million) [33, 37]. Furthermore, this same weighted least squares approach is easily extended to multiple meta-regression, where it again surpasses both fixed and random effects in much the same way as revealed here [30].

Acknowledgement

We acknowledge support from the Czech Science Foundation (grant 15-02411S).

References

1. Raudenbush SW. Random effects models in H. Cooper and L.V. Hedges (eds.) In *The Handbook of Research Synthesis*. Russell Sage: New York, 1994; 301–321.
2. Hedges LV. Fixed effects models in H. Cooper and L.V. Hedges (eds.) In *The Handbook of Research Synthesis*. Russell Sage: New York, 1994; 285–299.
3. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 1998; **3**:486–504.
4. Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine* 2010; **29**:2969–2983.
5. Stanley TD, Jarrell SB, Doucouliagos HC. Could it be better to discard 90% of the data? A statistical paradox. *The American Statistician* 2010; **64**:70–77.
6. Stanley TD, Doucouliagos CH. Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods* 2014; **5**:60–78.
7. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999; **150**:469–475.
8. Sutton AJ, Song F, Gilbody SM, Abrams KR. Modelling publication bias in meta-analysis: a review. *Statistical Methods in Medical Research* 2000; **9**:421–445.
9. Aitken AC. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh* 1935; **55**:42–48.
10. Davidson R, MacKinnon JG. *Econometric Theory and Methods*. Oxford University Press: Oxford, 2004.
11. Greene WH. *Econometric Analysis*. Macmillan: New York, 1990.
12. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, 2002.
13. Stanley TD, Jarrell SB. Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys* 1989; **3**:161–170.
14. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
15. Fazel S, Khosla V, Doll H, Geddes J. The prevalence of mental disorders among the homeless in Western countries: systematic review and meta-regression analysis. *PLOS Medicine* 2008; **5**(12):1670–1681.
16. Stanley TD. Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* 2008; **70**:103–127.
17. Baker WL, White CM, Cappelleri JC, Kluger J, Coleman CI. Understanding heterogeneity in meta-analysis: the role of meta-regression. *The International Journal of Clinical Practice* 2009; **63**(10):1426–1434.
18. Copas JB, Lozada C. The radial plot in meta analysis: approximations and applications. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2009; **58**:329–344.
19. Moreno SG, Sutton AJ, Ades A, Stanley TD, Abrams KR, Peters JL, Cooper NJ. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology* 2009; **9**(2). DOI: 10.1186/1471-2288-9-2.
20. Karkos CD, Sutton AJ, Bown MJ, Sayers RD. A meta-analysis and metaregression analysis of factors influencing mortality after endovascular repair of ruptured abdominal aortic aneurysms. *European Journal of Endovascular Surgery* 2011; **42**:775–786.
21. Drewes HW, Steuten LMG, Lemmens LC, Baan CA, Boshuizen HC, Elissen AMJ, Lemmens KMM, Meeuwissen JAC, Vrijhoef HJM. The effectiveness of chronic care management for heart failure: meta-regression analyses to explain the heterogeneity in outcomes. *HSR: Health Services Research* 2012; **47**(5):1926–1959.
22. Higgins JPT, Thompson SG. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* 2002; **21**:1539–1558.
23. Baker RD, Jackson D. Meta-analysis inside and outside of particle physics: two traditions that should converge? *Research Synthesis Methods* 2013; **4**:109–124.
24. DerSimonian R, Laird M. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
25. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* 2008; **358**:252–260.
26. Turner RM, Davy J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012; **41**: 818–27.
27. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Academic Press: Orlando, 1985.

28. Stead LF, Perera R, Bullen C, Mant D, Lancaster T. Nicotine replacement therapy for smoking cessation. *The Cochrane Library* 2008; (2). DOI: 10.1002/14651858.CD000146.pub4.
29. Rockers PC, Røttingen JA, Shemilt I, Tugwell P, Bärnighausen T. *Inclusion of quasi-experimental studies in systematic reviews of health systems research*. Mimeo: Harvard University, 2014.
30. Stanley TD, Doucouliagos CH. *Better than Random: Weighted Least Squares Meta-Regression Analysis. SWP*, Economics Series 2013/2. Deakin University: Melbourne, Australia. Available from: http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2013_2.pdf [Accessed on 16 March 2013].
31. Cohen J. *Statistical Power Analysis in the Behavioral Sciences*. Hillsdale: Erlbaum, 1988.
32. Doucouliagos CH, Stanley TD. Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations* 2009; **47**:406–429.
33. Doucouliagos CH, Stanley TD, Giles M. Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics* 2012; **31**:197–206.
34. Doucouliagos CH, Haman J, Stanley TD. Pay for performance and corporate governance reform. *Industrial Relations* 2012; **51**:670–703.
35. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial result. *Cochrane Review* 2009; **1**. DOI: 10.1002/14651858.MR000006.pub3.
36. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**:629–634.
37. Stanley TD, Doucouliagos CH. *Meta-Regression Analysis in Economics and Business*. Routledge: London, 2012.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.