Journal of the Korean Statistical Society
https://doi.org/10.1007/s42952-022-00169-1

Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

# Infinite diameter confidence sets in Hedges' publication bias model

**Jonas Moss**[1]

Received: 8 November 2021 / Accepted: 30 March 2022
© The Author(s) 2022

**Abstract**
Meta-analysis, the statistical analysis of results from separate studies, is a fundamental building block of science. But the assumptions of classical meta-analysis models are not satisfied whenever publication bias is present, which causes inconsistent parameter estimates. Hedges' selection function model takes publication bias into account, but estimating and inferring with this model is tough for some datasets. Using a generalized Gleser–Hwang theorem, we show there is no confidence set of guaranteed finite diameter for the parameters of Hedges' selection model. This result provides a partial explanation for why inference with Hedges' selection model is fraught with difficulties.

**Keywords** Meta-analysis · Confidence intervals · File-drawer problem · Publication bias · Selection models · Weight function models

## 1 Introduction

A meta-analysis is a statistical analysis that quantitatively combines results from separate scientific studies. When the studies measure the same phenomenon, pooling of information allows us to predict the common effect size with larger precision than we could have done with one study alone. Meta-analyses are ubiquitous in the empirical sciences and forms a key component of most systematic reviews such as Cochrane reviews (Higgins et al., 2019).

Most meta-analytic techniques assume honest and unbiased reporting of results. But there is ample evidence that the scientific literature is not unbiased, as studies with significant $p$-values tend to be published with a greater probability than other studies (Easterbrook et al., 1991), a phenomenon called publication bias by Sterling (1959) and the file-drawer problem by Rosenthal (1979). When publication bias is

✉ Jonas Moss
jonas.moss@bi.no

[1] Department of Data Science and Analytics, BI Norwegian Business School, Oslo, Norway

Published online: 22 April 2022

present, there is no reason to trust the results of meta-analytic methods that do not account for it, as the parameter estimates will be inconsistent (Carter et al., 2019).

Hedges' (1992) publication bias model takes publication bias explicitly into account using a selection model, and is arguably the most appropriate model for publication bias (Carter et al., 2019). Despite there being an R (R Core Team, 2020) package for maximum likelihood estimation of this model, called `weightr` (Coburn et al., 2019), the model has not yet taken off. Its maximum likelihood estimation methods are numerically unstable and its estimates can be off even when they converge (Coburn et al., 2019; Stanley, 2005). The estimate of the mean effect size may be negative and of unrealistically large magnitude, and the estimated heterogeneity parameter might be improbably large. It turns out there are ridges in the likelihood function that can be linked to this behavior (McShane et al., 2016), but it has not been stated in clear terms exactly what the consequences are for inferential procedures. The purpose of this note is to explain why Hedges' publication bias performs poorly, by showing there is no confidence set for the mean effect size that has infinite diameter with probability zero.

## 2 Hedges' publication bias model

The most popular and well-known meta-analysis method is the random effects model with normal likelihoods (Hedges and Vevea, 1998). Written in hierarchical notation, it equals

$$\begin{aligned}\theta_i &\sim N(\theta_0, \tau),\\ x_i \mid \theta_i, \sigma_i &\sim N(\theta_i, \sigma_i).\end{aligned}$$

Here $x_i$ is the effect size and $\sigma_i$ is the standard deviation of the $i$th study, $i = 1, \dots, N$. Following the convention in meta-analysis, we assume all $\sigma_i$s to be known. The mean parameter $\theta_0$ is the population effect size, $N(\theta_0, \tau)$ is the effect size distribution, and $\tau$ is the heterogeneity parameter. The purpose of the effect size distribution is to model the fact that most effect size estimates plugged into a meta-analysis do not appear to measure the same phenomenon. By integrating out $\theta_i$, we find the density of $x_i$,

$$f(x_i;\theta_0, \tau, \sigma_i) = \phi(x_i;\theta_0, (\tau^2 + \sigma_i^2)^{1/2}),$$

where $\phi$ is the density of a normal random variable.

We will assume that the random effects meta-analysis model is true in the absence of publication bias. The mechanisms that cause publication bias modify the density in a suitable way. Consider the case when only significant studies at some specified level $\alpha$ are published. Assuming one-sided tests, the $p$-values are $u_i = \Phi(-x_i/\sigma_i)$, or normal one-sided $p$-values. We will only deal with with one-sided $p$-values in this paper, as there is usually, but not always, just one direction that is interesting to researchers, reviewers, and editors. A one-sided $p$-value can also be used if the researchers reported a two-sided value, since $p = 0.05$ for a two-sided hypothesis corresponds to $p = 0.025$ for a one-sided hypothesis, et cetera.

Define $c_\alpha = \Phi^{-1}(1-\alpha)$, the cutoff for significance at level $\alpha$. The *basic publication bias model* is a truncated normal model with density

$$f(x_i;\theta_0,\sigma_i) = \Phi\left(\frac{\theta_0 - c_\alpha}{(\sigma_i^2+\tau^2)^{1/2}}\right)^{-1} \phi(x;\theta_0,(\sigma_i^2+\tau^2)^{1/2})1(x_i/\sigma_i > c_\alpha], \quad (2.1)$$

where $1[A]$ is the characteristic function of $A$. This model for publication bias was introduced by Hedges (1984) in the context of $F$-distributions.

The basic publication bias model is unrealistic. It requires that no non-significant studies are published. But even in the fields most severely affected by publication bias, such as psychology, a non-negligible number of non-significant studies are published (Motyl et al., 2017). Moreover, the basic publication bias model does not allow for different cutoffs for significance. It is likely that some editors will accept studies reaching a significance at $\alpha = 0.025$, corresponding to $x_i/\sigma_i > 1.96$ but not at $\alpha = 0.05$, corresponding to $x_i/\sigma_i > 1.64$.

These problems can be rectified by adopting the selection model for publication bias of Iyengar and Greenhouse (1988), which models the following scenario.

> **Publication bias scenario.** Alice the editor receives a study with the $p$-value $u$. Her publication decision is a random function of this $p$-value. That is, she will publish the study with some probability $w(u)$ and reject it with probability $1 - w(u)$. Every study you will ever read in Alice's journal has survived this selection mechanism, the rest are lost forever.

Let $w(u_i)$ be a function of the $p$-value $u_i = \Phi(-x_i/\sigma_i)$ taking values in [0, 1]. Then $w(u_i)$ is a probability for every $u_i$, and the selection model

$$f(x_i;\theta_0,(\sigma_i^2+\tau^2)^{1/2}) \propto \phi(x_i;\theta_0,(\tau^2+\sigma_i^2)^{1/2})w(u) \quad (2.2)$$

models the publication bias scenario exactly. This model can be viewed as a rejection sampling procedure (Flury, 1990; von Neumann, 1951), where $\phi$ serves as proposal distribution for $f$. Variants of this model, with and without covariates, has been studied by e.g. Dear and Begg (1992), Vevea and Hedges (1995), Vevea and Woods (2005), Citkowicz and Vevea (2017).

Hedges (1992) studies the selection model when $w$ is a step function with fixed steps. Let $\alpha$ be a vector with elements $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_K = 1$ and $\rho$ be a $K$-ary non-negative, non-increasing vector having its first element equal to $\rho_1 = 1$ for identifiability. Define the step function $w$ based on $\alpha$ and $\rho$ as

$$w(u;\rho,\alpha) = \sum_{k=1}^{K} \rho^k 1_{(\alpha_{k-1},\alpha_k]}(u). \quad (2.3)$$

We call the selection model with a step function *Hedges' publication bias model*. Its density is

$$f(x_i;\theta_0,(\tau^2+\sigma_i^2)^{1/2}) \propto \sum_{k=1}^{K} \rho^k \phi(x_i;\theta_0,(\tau^2+\sigma_i^2)^{1/2})1(\alpha_{k-1},\alpha_k](u_i). \tag{2.4}$$

Interpreting Hedges' publication bias model is easy. When the editor receives a study with $p$-value $u$, she finds the $k$ such that $u \in (\alpha_{k-1}, \alpha_k]$ and accepts with probability $\rho^k$. Since $\rho^1 = 1$, she always accepts when $u \in [0, \alpha_1]$. The vector $\rho$ is non-increasing since a publication decision based solely on $p$-values should always act favorably towards lower $p$-values. The parameters $(\theta_0, \tau, \rho)$ of the model are identified when $\alpha$ is fixed (Moss and De Bin, 2021, Web Appendix A).

It is probably not possible to generalize the results of this paper to selection models that do no follow the step function model. Lemma 5, about the truncated normal, is crucial in the proof of our main result, Theorem 8. Truncated densities only appear in step function models, not models with continuous selection functions, such as the Probit selection function of Copas (2013) or the one-parameter selection functions of Preston et al. (2004).

Hedges' publication bias model allows both for non-significant studies to be published and allows the editor to act differently towards different cutoffs such as $\alpha = 0.025$ and $\alpha = 0.05$. In addition, the model can approximate any non-increasing selection function $w$ by increasing the number of steps. In applications, the parameters $\mu$, $\tau$, and $\rho$ are estimated from the data, while $\alpha$ is fixed by the researcher, for instance at $\alpha = (0.025, 0.05, 1)$. We recommend using these cutoffs, as it is well known that applied journals frequently demand statistical significance at this level. Since both two-sided and one-sided $p$-values occur, we need to include 0.025 in addition to 0.05.

We can write Hedges' model as a mixture model on the form

$$f(x_i;\theta_0,(\tau^2+\sigma_i^2)^{1/2}) = \sum_{k=1}^{K} \pi^k f^k(x_i;\theta_0,(\tau^2+\sigma_i^2)^{1/2}). \tag{2.5}$$

where $f^k, k \le K$ are normal densities truncated to $(\Phi^{-1}(1-\alpha_{k-1}), \Phi^{-1}(1-\alpha_k)]$, and $\pi^k$ are mixture probabilities, i.e., $\pi^k > 0$ for each $k$ and $\sum_{k=1}^{K} \pi^k = 1$. The mixture probabilities $\pi^k$ are functions of $(\theta_0, \tau, \sigma_i, \rho)$, see the appendix (p. 10) for their formula.

The main benefit of Hedges' publication bias model (2.4) is how it models $p$-values based publication bias directly, there is no approximation involved. If you believe in the random effects meta-analysis model and the $p$-value based publication bias scenario, Hedges' publication bias model is simply the correct model. Most statistical methods correcting for publication bias in the literature either do not make use of an explicit statistical model or do not estimate the parameters $\theta_0$ and $\tau$. For instance, the funnel plot of Egger et al. (1997) is a graphical method, while the trim-and-fill method of Duval and Tweedie (2000) is a non-parametric method based on making the funnel plot symmetric. Stanley (2005); Stanley and Doucouliagos (2014) discuss various misspecified regression-based estimators of the corrected effect size $\theta$ based on the fixed effect Hedges' publication bias model. The estimating $p$-curve method of Simonsohn et al. (2014) and $p$-uniform of van Assen et al. (2015); van

Aert et al. (2016) are two methods for dealing with publication bias hailing from psychology. Both are based on a variant of the basic publication bias model, but with fixed instead of random effects, and both employ somewhat unusual estimation methods (McShane et al., 2016). Since there is ample evidence of heterogeneity in meta-analysis, restricting oneself to the fixed effects meta-analysis is a mistake.

Hedges' model has some downsides. It models only bias due to selection of $p$-values, not every source of bias, such as language bias (Egger and Smith, 1998). Second, it may not be best model for biases with other causes than the publication process, such as $p$-hacking (Simmons et al., 2011). Moss and De Bin (2021) propose a related model that may be more successful at correcting for $p$-hacking.

Hedges' model can be hard to estimate, especially if the data is unfavourable. Stanley (2005, Sect. 6.3) discusses three problematic cases in economics when maximum likelihood was used to estimate Hedges' publication bias model. McShane et al. (2016, Appendix A) notes that while estimation of the basic publication bias model is hard, introducing the heterogeneity parameter exacerbates the problem. The likelihood function has contours following approximately $\tau \propto |\theta|^{1/2}$.

Figure 1 shows the contour lines for the meta-analysis of Cuddy et al. (2018) where the selection probabilities of the step function (2.3) are fixed at $\rho = (1, 0.6, 0.1)$ and $\alpha = (0, 0.025, 0.05, 1)$. Only just around the maximum at $\hat{\theta}_0 = 0.55$ and $\hat{\sigma} = 5 \cdot 10^{-7}$ can the likelihood be approximated with a quadratic function.

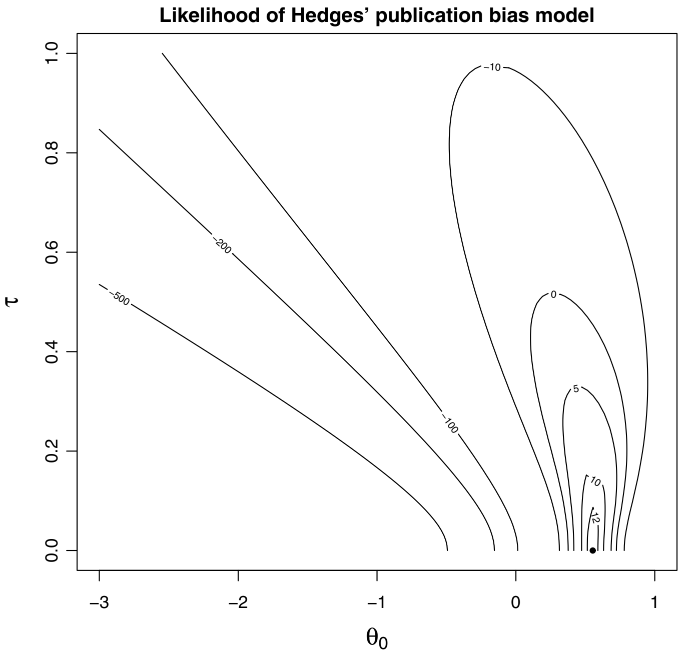


**Fig. 1** Contour lines for the log-likelihood for the simple publication model using power posing data of Cuddy et al. (2018)

Estimation of Hedges' model is especially hard when almost all observations lie close to the the cutoffs. In our experience, estimation works well when the data is sufficiently well spread, for instance when some observations are far away from the cutoffs and there are observations that have failed to reach significance. As a simple example of a case when estimation fails to work, consider the observation vector $x = (1.96, 1.96, 1.96, 1.96, 1.96, 1.64, 1.64, 1.64, 1.64, 1.64, -1)$ when the standard deviation of each observation is $\sigma_i = 1$. We employ $\alpha = (0, 0.025, 0.05, 1)$, which implies cutoffs at $\Phi^{-1}(0.975)$ and $\Phi^{-1}(0.95)$. In this case, 10 out of 11 observations are very close to the cutoffs. When we run the Hedges selection model on this data, using the `sel` function of the `R` package `metafor` Viechtbauer (2010), the model does not converge. If the last observation is changed to 0 instead of −1, the model converges, but the Hessian cannot be inverted, and the parameter estimate for $\theta_0$ equals −1.8. Situations similar to this one are most common when $n$ is small.

## 3 Confidence sets of infinite diameter

Fix some measurable space $(Ømega, \mathcal{F})$, let $\mathcal{P}$ be a family of dominated probability measures defined on this measurable space, and $\Pi$ a partition of $\mathcal{P}$. Recall that a partition of $\mathcal{P}$ is a collection of disjoint non-empty subsets $\pi$ of $\mathcal{P}$ such that $\bigcup_{\pi\in\Pi} \pi = \mathcal{P}$. When $p$ is a density associated with a $P \in \mathcal{P}$, we will use the standard notation $[p]$ to denote the unique part $\pi$ containing $P$. Instead of partitions, we could have used a formulation with main parameters $\theta$ and nuisance parameters $\eta$, and defined the rejection set for $\theta$ as $\sup_\eta \sup_\theta P_{\theta,\eta}(R(\theta)) \leq \alpha$. We have decided to use partitions for two reasons: First, there are no unambiguous nuisance and main parameters in our applications, making notation using nuisance parameters confusing. Second, the upcoming Theorem 4 can be applied in purely non-parametric situations, where the mention of nuisance and main parameters is even more confusing.

**Definition 1** A *confidence set* of level $\alpha$ is a family of rejection sets $\{R(\pi)\}, \pi \in \Pi$ such that

$$\sup_{\pi\in\Pi} \sup_{P\in\pi} P(R(\pi)) \leq \alpha.$$

If the inequality is an equality, the confidence set has *size* $\alpha$.

This definition of confidence sets might look slightly unfamiliar, but it is a straightforward generalization of the definitions in Casella and Berger (2002, Definition 9.1.5) and Lehmann and Romano (2005, Sect. 3.5). Usually, a confidence is defined as a set $C$ adhering to the relation

$$\pi_0 \in C \iff \omega \notin R(\pi_0). \tag{3.1}$$

That is, $\pi_0 \in C$ if and only if we accept the null-hypothesis $H_0 : \pi = \pi_0$, but we will only need the formulation using rejection sets in this paper. When confidence sets are defined in terms of rejection sets, there is sometimes no partition $\Pi$ to take the supremum over, and the definition reduces to $\sup_{P\in\mathcal{P}} P(R(P)) \leq \alpha$. The term *confidence interval* is far more common than confidence set, but this requires that the $C$ in equation 3.1 is an interval, which we will not require here.

The following example should make Definition 1 clearer.

***Example 2*** Consider the usual $t$-confidence interval with $n$ observations. In this case, $\mathcal{P}$ contains all measures $P^n_{\mu,\sigma}$, where $P_{\mu,\sigma}$ is the probability measure of a normal with mean $\mu$ and standard deviation $\sigma$. Here $Q^n$ denotes the $n$-fold product measure of $Q$, corresponding to $n$ independent samples from $Q$ when $Q$ is a probability measure. The $t$-confidence interval is an exact confidence interval for $\mu$ no matter what $\sigma > 0$, the nuisance parameter, is. Since the test is exact, rejection sets $R(\mu)$ satisfy $P^n_{\mu,\sigma}(R(\mu)) = \alpha$ for all $\sigma, \mu$. We can formulate the confidence set in terms of partitions too. Let $\pi(\mu) = \{P^n_{\mu,\sigma} \mid \sigma > 0\}$ contain all normal probability measures with mean $\mu$ and some positive standard deviation. Then $\{\pi(\mu)\}, \mu \in \mathbb{R}$ defines a partition of $\mathcal{P}$. Then the two-sided $t$-confidence interval is a confidence set of size $\alpha$ with partition $\Pi = \{\pi(\mu)\}, \mu \in \mathbb{R}$ according to Definition 1.

Now we must find out how to measure the size of confidence sets. To make our results as general as possible, we will let the *size function* be any non-negative function $\|\cdot\| : \Pi \to [0, \infty)$. In most cases, the size function will be a norm, but any non-negative function is a valid size function. For instance, in the $t$-confidence interval example above, $\|\cdot\|$ can be taken to be $\|\pi\| = |\mu|$ for the unique $\mu$ associated with each $\pi$.

The diameter of a confidence set is the random variable

$$D(\omega) = \sup_{\pi\in\Pi}\{\|\pi\| \mid \omega \notin R(\pi)\}. \tag{3.2}$$

The diameter tells you the size of the largest accepted $\pi$. We will assume that $D$ is Borel measurable.

**Definition 3** A confidence set has infinite diameter with $P$-positive probability if $P(D = \infty) > 0$. It has infinite diameter with positive probability if $P(D = \infty) > 0$ for all $P \in \mathcal{P}$.

The original Gleser–Hwang theorem (Gleser and Hwang, 1987, Theorem 1) is defined for pairs of parameters $\theta_1, \theta_2$, where $\theta_2$ is a nuisance parameter and the confidence set is constructed for a functional $\gamma(\theta_1)$. The following generalization does not require any nuisance parameters. Using partitions, it can be used both with and without nuisance parameters, as well as in non-parametric settings. Its proof is in the appendix (p. 9).

**Theorem 4** (Gleser–Hwang theorem) *Suppose there is a sequence* $\{p_n\}$ *of densities derived from* $\mathcal{P}$ *satisfying the following*:

(i) *There is a density* $p^\star$ *such that* $p_n$ *converges to* $p^\star$ *pointwise*,
(ii) $\operatorname{supp} p \supseteq \operatorname{supp} p^\star$ *for all densities p derived from* $\mathcal{P}$,
(iii) *the size of the equivalence class* $[p_n]$ *goes to infinity as n increases*, $\|[p_n]\| \to \infty$ .

Then every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

Following the terminology of Berger et al. (1999), we will say that families $\mathcal{P}$ of probabilities satisfying the conclusion of Theorem 4 for a suitable partition $\Pi$ belong to the Gleser–Hwang class. To make the Gleser–Hwang class more familiar, we will present two examples. More examples can be found in the papers of Gleser and Hwang (1987) and Berger et al. (1999).

Fieller's problem is the best known case of a badly behaved confidence set. Let $(X, Y)$ be an observation from a bivariate normal $N([\mu_1, \mu_2], I\sigma^2)$, where $\sigma^2$ is known. We want to form a confidence set for the ratio $\mu_2/\mu_1$. The most famous confidence set is due to Fieller (1940). His confidence set can be finite, the whole real line, or the union of two disjoint semi-infinite intervals, all with positive probability (Koschat, 1987).

But it is not only Fieller's confidence set that might be infinitely long. The Gleser–Hwang theorem can be used to show that *every* confidence set for $E(Y)/E(X)$ must be infinitely long with positive probability. This result is almost independent of the distribution of $X$ and $Y$. To state this result in our notation, let $\mathcal{P}$ be a family of bivariate distributions over $(X, Y)$. All of these distributions have the same support, and all of them have finite means $E_p(X)$ and $E_p(Y)$. Moreover, assume $E_p(X) = 0$ is attainable for some $p \in \mathcal{P}$. Define the partition $\Pi$ by $p, q \in \pi$ if and only if $E_p(X)/E_p(Y) = E_q(X)/E_q(Y)$, and let $\|[p]\| = |E_q(X)/E_q(Y)|$, i.e., the ratio of means. Choose a sequence $p_n(x, y) = p(x, y)$, where $p(x, y)$ is density with means $E_pX > 0$ and $E_pY = 0$. Then $\|[p]\| = \infty$, the conditions of Theorem 4 are satisfied, and every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

Another example is due to Bahadur and Savage (1956), who studies non-parametric testing of the mean, and concludes the mean cannot be meaningfully tested. They are working with a family $\mathcal{P}$ of densities over $\mathbb{R}$ that covers all finite means, has finite variances, and is closed under convex combinations. Similar problems were considered by Romano (2004) and Donoho (1988).

Using the Gleser–Hwang theorem, it is easy to verify that every confidence set has infinite diameter with positive probability. Define the partition $\Pi$ by $p, q \in \pi$ if and only if $E_p(X) = E_q(X)$, and let $\|[p]\| = |E_p(X)|$. Let

$$p_n(x) = \left(1 - \frac{1}{n}\right) q_0(x) + \frac{1}{n} q_{n^2}(x),$$

where $q_0$ has mean 0 and $q_{n^2}$ has mean $n^2$. Then $\|[p_n]\| = n$, the conditions of Theorem 4 are satisfied, and every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

There are several natural candidates for $\Pi$ when working with Hedges' selection function model (2.4). We will work with three of them. First, consider the partition where $p, q \in \pi$ if and only if $p$, $q$ have the same mean effect size parameter $\theta_0$. We will equip this partition with the size function $\|\pi\| = |\theta_0|$, and it corresponds to a confidence set for $\theta_0$. Second, consider the partition where all $p, q \in \pi$ have the same heterogeneity parameter $\tau$, equipped with $\|\pi\| = \tau$. Finally, we will work with the partition where all $p, q \in \pi$ have the same heterogeneity parameter $\tau$ and population effect size $\theta_0$, and equip it with $\|\pi\| = (\theta_0^2 + \tau^2)^{1/2}$. This information is summarized in Table 1 for convenience.

Let us take a look at the basic publication bias model (2.1) again. To use Theorem 4 we need a witnessing sequence of functions $p_n \to p$ satisfying the conditions (ii) and (iii). The next lemma shows how to make such a witness for the truncated normal. Its proof is in the appendix (p. 9).

**Lemma 5** *Let $f_n$ be a normal density truncated to $[a, b)$, where $b = \infty$ is allowed, with underlying mean $\theta_n = -n$ and standard deviation $\sigma_n^2 = n + c$ for some $c \in \mathbb{R}$. Then $f_n$ converges pointwise to* $\exp(-x)/[\exp(-a) - \exp(-b)]$, *the distribution of an exponential variable truncated to* $[a, b)$.

Using Lemma 5 it is not hard to show that the basic publication bias model (2.1) is a member of the Gleser–Hwang class.

**Theorem 6** *Assume we have N independent samples from the basic publication bias model* (2.1). *Then any confidence set for $\theta_0, \tau$, or $(\theta_0, \tau)$ with level $\alpha > 0$ will have infinite diameter with positive probability.*

***Proof*** Let $\Pi$ be the partition of $\mathcal{P}$ where $p, q \in \pi$ if and only if they share the same $\theta_0$, and let $||[p]|| = |\theta_0|$. We are dealing with products of densities of the form (2.1), that is,

$$p(x) = \prod_{i=1}^{N} \Phi\left(\frac{\theta_0 - c_\alpha}{(\sigma_i + \tau)^{1/2}}\right)^{-1} \phi(x_i; \theta_0, (\sigma_i^2 + \tau^2)^{1/2}),$$

where $\sigma_i$ are known parameters. From Lemma 5, $p_n$ converges to a product of truncated exponentials when $\theta_n = -n$ and $\tau_n^2 = n$. Since $||[p_n]|| = n$, the three conditions of Theorem 4 are satisfied. The proofs for $||[p]|| = \tau$ and $||[p]|| = (\tau^2 + \theta_0^2)^{1/2}$ are similar and omitted. □

**Table 1** The three partitions $\Pi$ for the selection function model

| | Symbol | Size $\|\cdot\|$ | Confidence set |
|---|---|---|---|
| Mean effect size | $\theta_0$ | $\|\pi\| = \|\theta_0\|$ | Confidence set for $\theta_0$ |
| Heterogeneity parameter | $\tau$ | $\|\pi\| = \tau$ | Confidence set for $\tau$ |
| Both parameters | $(\theta_0, \tau)$ | $\|\pi\| = (\theta_0^2 + \tau^2)^{1/2}$ | Joint confidence set for $(\theta, \tau)$ |

Proving the analogue of Theorem 6 for Hedges' publication is only somewhat more involved. We will use the mixture representation of (2.5) and a lemma generalizing Theorem 4 to a certain kind of mixtures.

Let $f^1$,$f^2$, …,$f^K$ be a sequence of densities, $\pi = (\pi^1, \pi^2, \dots, \pi^K)$ be a probability vector, and $p = \sum_{k \le K} \pi^k f^k$ be a mixture distribution. We will assume that the size of $[p]$ equals the size of any of its mixture components $[f^k]$ for some size $\|\cdot\|$, i.e., $\|[p]\| = \left\|[f^k]\right\|$ for all $k$. Why we do this will be clear in the proof of Theorem 8, but think of it this way: If all of $p$s mixture components have the same mean, the mean of $p$ equals the mean of any $f^k$.

**Lemma 7** *Let $\mathcal{P}$ be a class of K-ary mixture distributions and $||[p]||$ be as assumed above. Assume there is a sequence $p_n = \sum_{k \le K} \pi_n^k f_n^k$ and a subset $K'$ such that*

(i) *For all $k \in K'$, there is a density $f^{k\star}$ such that $f_n^k$ converges to $f^{k\star}$ pointwise.*
(ii) *For all mixtures $p$, $\operatorname{supp} p \supseteq \operatorname{supp} f^{k\star}$ for all $k \in K'$.*
(iii) *For all $k \in K'$, the size of $[f_n^k]$ goes to infinity, $\left\|[f_n^k]\right\| \to \infty$.*
(iv) *The density concentrates on the components indexed by $K'$, $\lim_{n\to\infty} \sum_{k\in K'} \pi_n^k = 1$.*

Then every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

***Proof*** We employ Theorem 4. By (i) and (iv), $p_n$ converges pointwise to the density

$$\sum_{k\in K'} \pi_n^{k\star} f_n^{k\star} = p^\star.$$

That $\operatorname{supp} p \supseteq p^\star$ follows from (ii) and (iv). Finally, from the assumption that $\left\|[p_n]\right\| = \left\|[f_n^k]\right\|$, we get that $\left\|[p_n]\right\| \to \infty$ too. □

**Theorem 8** *Assume we have N independent samples from the publication bias model (2.1), where the selection probability $\rho$ is unknown and $\alpha$ is known. Then any confidence set for $\theta_0, \tau$, or $(\theta_0, \tau)$ will have infinite diameter with positive probability.*

***Proof*** Let $n = 1$ and consider confidence sets for $\theta_0$. Let $\Pi$ be the partition of $\mathcal{P}$ where $p, q \in \pi$ if and only if they share the same $\theta_0$, and let $||[p]|| = |\theta_0|$. Then $||f^k|| = |\theta_0|$ from the mixture representation (2.5). Using Lemma 5, we see that $f^k, k < K$ converges pointwise to truncated exponentials when $\theta_0 = -n$ and $\tau_n^2 = n$, so that (i), (iii) of Proposition 7 are satisfied with the set $K' = \{1, 2, \dots, K - 1\}$. Moreover, since we assume that $\rho$ is decreasing, (ii) is satisfied as well. The mixture probabilities for $k \ne K$ can be fixed at e.g. $\pi = 1/(K - 1)$, and (iv) is satisfied as well. The proofs for $||[p]|| = \tau$ and $||[p]|| = (\tau^2 + \theta_0^2)^{1/2}$ are similar and omitted.

When $N > 1$, expand the expression $\prod_{i=1}^N \sum_{k<K} \pi_i^k(\sigma_i) f_i^k(\sigma_i)$, and use the same reasoning as in the first part of this proof. □

## 4 Remarks

Well-behaved confidence sets for Hedges publication bias model do not exist, but well-behaved credibility sets do. Bayesian estimation of Hedges' model can be made routine, as it is easy to find uncontroversial priors for $\theta_0$ and $\tau$. In practical meta-analyses we know that $\theta_0$ cannot be large, and is likely to be close to 0. Moreover, since it is common effects in meta-analyses to be interpreted as the aggregation of many small effects, the central limit theorem justifies using a normal prior. As we want to remove prior mass from negative $\theta_0$s of large magnitude, $N(0, 1)$ is a decent standard prior. Similarly, a half-normal is a reasonable prior for the heterogeneity parameter $\tau$. Moss and De Bin (2021) employed these priors on several examples.

## Appendix

The following sandwich convergence theorem is used in the proof of the Gleser–Hwang theorem.

**Lemma 9** (Billingsley (1995, Exercise 16.4(a))) *Suppose the functions $a_n, b_n, f_n$ converge pointwise to $a$, $b$, $f$ and $a_n \leq f_n \leq b_n$ for all $n$. If $\int a_n d\mu \to \int a d\mu$ and $\int b_n d\mu \to \int b d\mu$, then $\int f_n d\mu \to \int f d\mu$ for any measure $\mu$.*

The proof of Theorem 4 closely follows the proof of Gleser and Hwang (1987, Theorem 1).

***Proof of Theorem 4*** We can assume without loss of generality that $\|[p_n]\| \geq n$, as we can choose a suitable sub-sequence if we have to. By definition of the diameter $D$ (3.2) we see that

$$\{D \geq n\} = \{\omega \in Ømega \mid \text{there is a } \pi \text{ such that} \|\pi\| \geq n \text{ and } \omega \in R^c(\pi)\}.$$

It follows that, if $||[p_n]|| \geq n$, then $R^c([p_n]) \subseteq \{D \geq n\}$. Since we assume that $\|[p_n]\| \geq n$ and

$$1 - \alpha \leq P_n(R^c([p_n])) = \int_{R^c([p_n])} p_n d\mu$$

by definition of a confidence set, we have that

$$0 < 1 - \alpha \leq \int_{R^c([p_n])} p_n d\mu \leq \int_{D \geq n} p_n d\mu \tag{4.1}$$

for all $n$. Since $p_n$ and $p^\star$ are densities,

$$\lim_{n\to\infty}\int p_n d\mu = 1 = \int p^\star d\mu = \int \lim_{n\to\infty} p_n d\mu.$$

This allows us to use Lemma 9 with $a_n = 0$, $b_n = p_n$, and $f_n = 1_{D\geq n}p_n$ and conclude that

$$\int_{D\geq n} p_n d\mu \to \int_{D=\infty} p^\star d\mu. \tag{4.2}$$

Combining equations (4.1) and (4.2), we get

$$0 < 1-\alpha \leq \int_{D=\infty} p^\star d\mu.$$

Let $P \in \mathcal{P}$ be arbitrary, $p$ be its density, and consider

$$P(D=\infty) = \int_{D=\infty} p d\mu \geq \int_{D=\infty\cap \operatorname{supp} p^\star} \left(\frac{p}{p^\star}\right) p^\star d\mu.$$

Since $\int_{D=\infty} p^\star d\mu > 0$ and $p/p^\star > 0$ on $\operatorname{supp} p^\star$ (since $\operatorname{supp} p \supseteq \operatorname{supp} p^\star$ by assumption), we see that $\int_{D=\infty\cap \operatorname{supp} p^\star}(p/p^\star)p^\star d\mu > 0$ too. It follows that $P(D=\infty) > 0$, and, since $P$ is arbitrary, $D$ has infinite diameter with positive probability. □

Now we prove Lemma 5.

***Proof of Lemma (5)*** Let $n > -c$, so that $\sigma_n^2 > 0$. Recall the well-known formula for the normal truncated to $[a, b]$, and substitute $\theta_n = -n$ and $\sigma_n^2 = n + c$,

$$\begin{aligned} f_n(x) =& \frac{1}{\Phi\left(\frac{b-\theta_n}{\sigma_n}\right) - \Phi\left(\frac{a-\theta_n}{\sigma_n}\right)} \phi(x;\theta_n,\sigma_n)1[a,b](x), \\ =& \frac{\phi(x;-n,(n+c)^{1/2})1[a,b](x)}{\Phi[-(a+n)(n+c)^{-1/2}] - \Phi[-(b+n)(n+c)^{-1/2}]}. \end{aligned} \tag{4.3}$$

The normal density part equals

$$\phi(x;-n,(n+c)^{1/2}) = (2\pi)^{-1/2}(n+c)^{-1/2}\exp\left(-\frac{x^2+2nx+n^2}{2(n+c)}\right).$$

When $n$ is large compared to $x$, the term $x^2/2(n+c)$ is negligible, hence

$$\begin{aligned} \phi(x;-n,(n+c)^{1/2}) \approx& (2\pi)^{-1/2}(n+c)^{-1/2}\exp(-n^2/2(n+c))\exp(-x), \\ =& (n+c)^{-1/2}\phi(n/(n+c)^{1/2})\exp(-x). \end{aligned}$$

From Equation 5 of Borjesson and Sundberg (1979), we know that $\Phi(-x) \approx \phi(x)/x$ as $x$ grows. Then

$$\Phi[-(a+n)(n+c)^{-1/2}] \approx \frac{(n+c)^{1/2}}{n+a}\phi[-(a+n)(n+c)^{-1/2}],$$

and using the same reasoning as above, we find that $\phi[-(a+n)(n+c)^{-1/2}] \approx \phi((n+c)^{1/2})\exp(-a)$ as $n$ increase. Therefore,

$$\Phi[-(a+n)(n+c)^{-1/2}] \to \frac{(n+c)^{1/2}\phi((n+c)^{1/2})\exp(-a)}{a+n}.$$

Since this reasoning applies to $b$ as well, we get that $f$ approaches

$$\begin{aligned}
&\frac{(n+c)^{-1/2}\phi(n/(n+c)^{1/2})\exp(-x)}{\Phi[-(a+n)n^{-1/2}] - \Phi[-(b+n)n^{-1/2}]},\\
&\approx \frac{(n+c)^{-1/2}\phi(n/(n+c)^{1/2})\exp(-x)}{(n+c)^{1/2}\phi((n+c)^{1/2})\left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]},\\
&= \frac{\phi(-n^2/2(n+c))\exp(-x)}{\phi((n+c)^{1/2})\left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]},\\
&\approx \exp(-x)n^{-1}\left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]^{-1},\\
&\to \exp(-x)\left[\exp(-a) - \exp(-b)\right]^{-1}.
\end{aligned}$$

Here the third line follows from $\phi(n/(n+c)^{1/2})/\phi((n+c)^{1/2}) \to 1$. □

These are the formula for the mixture probabilities $\pi_i$, see (2.5). Let $c_k = \Phi^{-1}(1-\alpha_k)$ and define

$$c = \sum_{k=1}^{K} \rho^k[\Phi(c_{k-1};\theta_0,(\tau^2+\sigma_i^2)^{1/2}) - \Phi(c_k;\theta_0,(\tau^2+\sigma_i^2)^{1/2})].$$

Then

$$\pi^k = c^{-1}\rho^k[\Phi(c_{k-1};\theta_0,(\tau^2+\sigma_i^2)^{1/2}) - \Phi(c_k;\theta_0,(\tau^2+\sigma_i^2)^{1/2})].$$

**Funding** Open access funding provided by Norwegian Business School.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visithttp://creativecommons.org/licenses/by/4.0/.

## References


Bahadur, R. R., & Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics, 27*(4), 1115–1122. https://doi.org/10.1214/aoms/1177728077

Berger, J. O., Liseo, B., & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science, 14*(1), 1–28. https://doi.org/10.1214/ss/1009211804

Billingsley, P. (1995). *Probability and measure*. USA: John Wiley & Sons.

Borjesson, P., & Sundberg, C. (1979). Simple approximations of the error function Q(x) for communications applications. *IEEE Transactions on Communications, 27*(3), 639–643. https://doi.org/10.1109/TCOM.1979.1094433

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. USA: Duxbury Press.

Citkowicz, M., & Vevea, J. L. (2017). A parsimonious weight function for modeling publication bias. *Psychological Methods, 22*(1), 28–41. https://doi.org/10.1037/met0000119

Coburn, K. M., Vevea, J. L., & Coburn. (2019). R package ‘weightr’ [Computer software manual]. cran.rstudio.org. Retrieved from https://cran.rstudio.org/web/packages/weightr/weightr.pdf

Copas, J. B. (2013). A likelihood-based sensitivity analysis for publication bias in meta-analysis. *Journal of the Royal Statistical Society. Series C, Applied statistics, 62*(1), 47–66. https://doi.org/10.1111/j.1467-9876.2012.01049.x

Cuddy, A. J. C., Schultz, S. J., & Fosse, N. E. (2018). P-Curving a more comprehensive body of research on postural feedback reveals clear evidential value for power-posing effects: Reply to Simmons and Simonsohn (2017). *Psychological Science, 29*(4), 656–666. https://doi.org/10.1177/0956797617746749

Dear, K. B. G., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a Meta-analysis. *Statistical Science, 7*(2), 237–245. https://doi.org/10.1214/ss/1177011363

Donoho, D. L. (1988). One-sided inference about functionals of a density. *Annals of Statistics, 16*(4), 1390–1420. https://doi.org/10.1214/aos/1176351045

Duval, S., & Tweedie, R. (2000). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867–872. https://doi.org/10.1016/0140-6736(91)90201-y

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Egger, M., & Smith, G. D. (1998). Bias in location and selection of studies. *British Medical Journal, 316*(7124), 61–66. https://doi.org/10.1136/bmj.316.7124.61

Fieller, E. C. (1940). The biological standardization of insulin. *Supplement to the Journal of the Royal Statistical Society, 7*(1), 1–64. https://doi.org/10.2307/2983630

Flury, B. D. (1990). Acceptance-rejection sampling made easy. *SIAM Review, 32*(3), 474–476. https://doi.org/10.1137/1032082

Gleser, L. J., & Hwang, J. T. (1987). The nonexistence of 100(1 - $\alpha$)% confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics, 15*(4), 1351–1362. https://doi.org/10.1214/aos/1176350597

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational and Behavioral Statistics, 9*(1), 61–85. https://doi.org/10.3102/10769986009001061

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science, 7*(2), 246–255. https://doi.org/10.1214/ss/1177011364

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486. https://doi.org/10.1037/1082-989x.3.4.486

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. USA: John Wiley & Sons.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science, 3*(1), 109–117. https://doi.org/10.1214/ss/1177013012

Koschat, M. A. (1987). A characterization of the fieller solution. *Annals of Statistics, 15*(1), 462–468. https://doi.org/10.1214/aos/1176350282

Lehmann, E. L., & Romano, J. P. (2005). Testing Statistical Hypotheses. Springer Science & Business Media. https://doi.org/10.1007/0-387-27605-x

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in Meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Moss, J., & De Bin, R. (2021, September). Modelling publication bias and p-hacking. *Biometrics*. Retrieved from https://doi.org/10.1111/biom.13560

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., & Skitka, L. J. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology, 113*(1), 34–58. https://doi.org/10.1037/pspa0000084

Preston, C., Ashby, D., & Smyth, R. (2004). Adjusting for publication bias: modelling the selection process. *Journal of Evaluation in Clinical Practice, 10*(2), 313–322. https://doi.org/10.1111/j.1365-2753.2003.00457.x

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna: Austria.

Romano, J. P. (2004). On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scandinavian Journal of Statistics, 31*(4), 567–584. https://doi.org/10.1111/j.1467-9469.2004.00407.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys, 19*(3), 309–345. https://doi.org/10.1111/j.0950-0804.2005.00250.x

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. *Journal of the American Statistical Association, 54*(285), 30–34. https://doi.org/10.1080/01621459.1959.10501497

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*(5), 713–729. https://doi.org/10.1177/1745691616650874

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods, 20*(3), 293–309. https://doi.org/10.1037/met0000025

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of bublication bias. *Psychometrika, 60*(3), 419–435. https://doi.org/10.1007/bf02294384

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

von Neumann, J. (1951). Various techniques used in connection with random digits. *Applied Math Series, 12*(36–38), 5.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.