



Universidad Autónoma
de Madrid

Biblos-e Archivo
Repositorio Institucional UAM

Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Psychological Methods 28.4 (2023): 895–904

DOI: <https://doi.org/10.1037/met0000446>

Copyright: © 2024 Asociación Estadounidense de Psicología

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

©American Psychological Association, 2021. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: <https://doi.org/10.1037/met0000446>

METHODS FOR ESTIMATING THE SAMPLING VARIANCE OF THE STANDARDIZED MEAN DIFFERENCE

Manuel Suero¹, Juan Botella¹, & Juan I. Durán²

¹ Universidad Autónoma de Madrid

² Universidad a Distancia de Madrid

Financial support: Ministerio de Ciencia e Innovación of Spain (project reference: PSI2017-82490-P).

Authors' Note:

This research was supported by a grant from the Ministerio de Economía, Industria y Competitividad of Spain, Project No. PSI2017-82490. The main ideas and results of this article were presented at the 9th European Congress of Methodology, held in Valencia, Spain, July 21-23, 2021.

Correspondence concerning this article should be addressed to Manuel Suero, Facultad de Psicología, Universidad Autónoma de Madrid, Campus de Cantoblanco, c/Ivan Pavlov, 6, 28049 Madrid, Spain. E-mail: manuel.suero@uam.es

Abstract

One of the most widely used effect size indices for meta-analysis in psychology is the *standardized mean difference* (*SMD*). The most common way to synthesize a set of estimates of the *SMD* is to weight them by the inverse of their variances. For this, it is necessary to estimate the corresponding sampling variances. Meta-analysts have a formula for obtaining unbiased estimates of sampling variances, but they often use a variety of alternative, simpler methods. The bias and efficiency of five different methods that have been proposed and that are implemented in different computerized calculation tools are compared and assessed. The data from a set of published meta-analyses are also re-analyzed, calculating the combined estimates and their confidence intervals, as well as estimates of the specific, between-studies variance, using the five estimation methods. This test of sensitivity shows that the results of a meta-analysis can change noticeably depending on the method used to estimate the sampling variance of *SMD* values, especially under a random-effects model. Some practical recommendations are made about how to choose and implement the methods in calculation resources.

Keywords: Standardized Mean Difference, Effect Size Variance, Sampling Variance of *d*, Sampling Variance of *g*, Meta-analysis.

METHODS FOR ESTIMATING THE SAMPLING VARIANCE OF THE STANDARDIZED MEAN DIFFERENCE

One of the main goals in meta-analysis is to obtain a pooled estimate of the effect size (*ES*) from a set of studies, each of which contributes an independent estimate. The combined estimate is obtained by weighting these estimates; the most frequently used weighting scheme involves the inverse of the variances of the estimators. Let us assume an estimator g of the parameter δ , or *standardized mean difference (SMD)*. We have k estimates and each one has its own sampling variance, σ_i^2 . The weight of each study is defined as $w_i = 1/\sigma_i^2$. Then, the combined estimate is obtained by

$$g_{\bullet} = \frac{\sum_{i=1}^k w_i \cdot g_i}{\sum_{i=1}^k w_i} \quad (1)$$

Of course, the weights w_i are unknown because the variances σ_i^2 are unknown. In practical work, the variances are replaced by their estimates, $\hat{\sigma}_i^2$, and with them the estimates of the weights, $\hat{w} = 1/\hat{\sigma}_i^2$, can be obtained.

In meta-analysis, the distinction between the fixed effect model and the random effects model is important (Borenstein, Hedges, Higgins, & Rothstein, 2010), although in practice the random effects model is almost always considered more appropriate. In the fixed effect model, all the studies share a single common *ES*, so that the variations observed in the estimates are due to the sampling of individuals. This is the sampling variance that is the focus of this article. Since the k studies have different sample sizes, each one has its own sampling variance, σ_i^2 . This is the only variance involved in the

weights in (1). The random effects model assumes a parametric distribution of *ES*s with a between-studies or specific variance, usually represented as τ^2 . This quantity represents the degree to which the initial conditions in the study and the procedures employed result in variations in the true, parametric *ES*. The sampling of true parametric effects constitutes a second source of heterogeneity. In the end, the variations observed in the estimates under a random effects model are due to the heterogeneity of the parametric values plus the sampling variance around the parametric value of each study. While in the fixed effect model the weights are $w_i = 1/\sigma_i^2$, in the random effects model they are $w_i = 1/(\sigma_i^2 + \tau^2)$. In this article we focus on the sampling variances, σ_i^2 , and specifically on their estimation given a specific parametric value. This parametric value may be the only one involved in the meta-analysis (fixed effect model) or the specific value of each study (random effects model). Later we will see that in the random effects model the method of estimating the sampling variance has consequences in the estimation of the other component of the variation, τ^2 .

As far as we know, five methods to obtain estimates of the sampling variance of the *SMD* have been proposed. In the present study, those five methods are assessed and compared. There are several reasons why it is important to obtain unbiased and accurate estimates of the sampling variances of the *ES* estimator. The most obvious is that the pooled estimate will be a better estimate of the parameter as the estimates of the variances of the studies are improved. But there are other reasons, as the quality of the estimates has consequences on other elements of the meta-analysis. For example, the variance of the combined estimator is estimated through $\hat{\sigma}^2(g_{\bullet}) = 1/\sum \hat{w}_i$. If the variance of g is estimated with a formula that provides over-estimated or underestimated values, then the weights will also be underestimated or over-estimated and the standard error of the combined estimator will be incorrect. Consequently, errors can occur both

in the significance tests (i.e., type I and II error rates could deviate from their nominal values) and in the confidence intervals (the coverage could deviate from the nominal level of confidence).

On the other hand, the quality of the estimates of the sampling variances will have additional consequences when fitting random effects models. The total empirical variance is obtained through the sum of squares of the individual estimates with respect to their combined estimate (Hedges & Olkin, 1985). Then, the specific variance, τ^2 , is estimated by decomposing the empirical variance and assuming the sampling variance as one of its components. If the sampling variances of the primary studies are over-estimated, then τ^2 will be underestimated, and vice versa.

Advancing our main conclusion, the differences between the several methods to estimate the sampling variance that we compare here are small at the study level, but their consequences at the meta-analysis level are not negligible (see White & Thomas, 2005), especially under a random effects model. Many sources recommend an approximate, but simple, formula for estimating the *SMD* sampling variance, highlighting that it is suitable for "large samples". The boundary between what are large samples and what are not is arbitrary and a clear operationalization is rarely proposed. Furthermore, very often the meta-analyst finds for the meta-analysis some primary studies with samples that would be considered operationally "large" and some primary studies with what would be considered "small" samples. It would be inconsistent to use different estimation procedures for the primary studies of the same meta-analysis. Therefore, it is necessary to establish unique rules that are suitable for meta-analyses that include primary studies with heterogeneous sample sizes.

We will first conduct a theoretical analysis and compare the amount of bias of the five estimators and their variances at the study level. We will also highlight practical

tools in which each of these estimators are implemented. Then we will perform a sensitivity test, fitting random-effects models to the data bases of a number of published meta-analyses. We will see that the results change significantly depending on the method chosen. We will conclude with some practical recommendations on choosing a method to estimate the sampling variance of the *SMD*.

The sampling variance of the *SMD*

Suppose two populations with distributions $X_i \sim N(\mu_1, \sigma)$ and $X_i \sim N(\mu_2, \sigma)$; that is, we assume that they are normal and homoscedastic. The *SMD* is defined as:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}. \quad (2)$$

In a primary study, a random sample of n_1 observations is drawn from the first population and a random sample of n_2 observations from the second. The mean and variance are calculated for each sample. The two estimates of the variances allow obtaining a joint estimate of the common variance, S_{pooled}^2 , so that the parameter δ can be estimated by,

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}. \quad (3)$$

It is well known that d is a biased estimator of δ . The bias is smaller as the sample sizes, n_1 and n_2 , increase, but with small samples it reaches a size that is not inconsequential. Hedges (1981) proposed an effective correction:

$$c(m) \cong 1 - \frac{3}{4 \cdot m - 1}, \quad (4)$$

where m are the degrees of freedom of the corresponding t test ($m = n_1 + n_2 - 2$)¹. Then, the unbiased estimate of δ is

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}} \cdot c(m) = d \cdot c(m). \quad (5)$$

From (5) it follows that the variances of g and d are closely related:

$$Var(g) = [c(m)]^2 \cdot Var(d). \quad (6)$$

As $0 < c(m) < 1$, necessarily $Var(g) < Var(d)$.

Hedges (1981, 1983) derived the formula for the parametric variance of g ,

$$\sigma^2(g) = \frac{a}{\tilde{n}} \cdot (1 + \tilde{n} \cdot \delta^2) - \delta^2. \quad (7)$$

where $a = [c(m)]^2 \cdot m / (m - 2)$ and $\tilde{n} = (n_1 n_2) / (n_1 + n_2)$. Formula (7) gives the parametric variance of g . The formula for the parametric variance of d , as presented by Hedges and Olkin (1985) is shown in the appendix, although it can be calculated through formulas

¹ The exact formula of the correction factor is $c(m) = \frac{\Gamma(m/2)}{\sqrt{m/2} \cdot \Gamma[(m-1)/2]}$, where $\Gamma(x)$ is the gamma

function (Hedges, 1981).

(7) and (6). Hedges and Olkin's book (1985) synthesizes earlier conclusions and sets the foundations of statistical analysis in meta-analysis. They also present an approximate formula for the variance of d , that is easier to apply with a pocket calculator. It is important to pay special attention to the terminology, since in that source what is called d in many other sources (and here) is called g , and vice versa. For example, in the formula on page 80 of Hedges and Olkin (1985), " $Var(g)$ " appears, but g is defined on page 78 as the biased estimator (3), which here we have called d . In the present terminology the approximate formula proposed for the variance of d is (Hedges & Olkin, 1985):

$$\sigma^2(d) \cong \frac{1}{\tilde{n}} + \delta^2 \cdot \frac{1}{2 \cdot (N - 3.94)} . \quad (8)$$

Several estimators of the sampling variance of g

To obtain the parametric variance of the unbiased estimator, g , with formula (7), or the approximate variance of the biased estimator, d , with (8), it is necessary to know the value of the parameter, δ , and this is normally unknown. We will compare five methods to estimate the variance of the unbiased estimator, $\hat{\sigma}^2(g)$, when δ is unknown. As we do not know of any other methods, we consider this set to be exhaustive. We will start by showing a common structure for the five estimators that will facilitate their comparison and then we will use this structure to derive the expected value and the variance of the five estimators. The common structure or *general representation* of the estimators evaluated here is:

$$\hat{\sigma}^2(g) = K1(n_1, n_2) + K2(n_1, n_2) \cdot d^2, \quad (9)$$

where d has been defined in (3), and $K1(n_1, n_2)$ and $K2(n_1, n_2)$ are expressions that depend only on the samples sizes². We will show that the estimators evaluated here differ in their ways of specifying $K1$ and $K2$. We will denote the different estimators with the subscript j in $\hat{\sigma}_j^2(g)$.

Unbiased estimator of Hedges (1983).

Hedges (1983) proposed the following estimator of the variance of g and showed that it is an unbiased estimator:

$$\hat{\sigma}_1^2(g) = \frac{1}{\tilde{n}} + \left(1 - \frac{1}{a}\right) \cdot g^2. \quad (10)$$

Given (5), substituting g^2 we arrive at the values of $K1$ and $K2$ for this estimator, in terms of the general expression (9):

$$\hat{\sigma}_1^2(g) = \frac{1}{\tilde{n}} + \left(1 - \frac{1}{a}\right) \cdot [c(m)]^2 \cdot d^2. \quad (11)$$

The estimator $\hat{\sigma}_1^2(g)$ is implemented, for example, in *metafor* (Viechtbauer, 2010).

This *R* package uses the "escalc" function to calculate estimates of the *ES* and its

² In order to simplify the nomenclature, from here on we will only write $K1$ and $K2$, omitting n_1 and n_2 ; it is implicit that the values of $K1$ and $K2$ actually are functions of the sample sizes.

variances. When the sample statistics are entered and *SMD* is chosen, it returns the value of g according to formula (5). A choice is offered between three estimators of the variance using the argument “vtype”. The option "UB" provides the estimator $\hat{\sigma}_1^2(g)$.

Approximate formula of Hedges and Olkin (1985).

Hedges and Olkin (1985) arrived at an approximate formula for estimating the variance of g , that is easier to apply than formula (10) with a pocket calculator. They recommended its use with large samples, assuming asymptotic normality (remember again the change in terminology, since this formula, which appears on page 86 of Hedges and Olkin, 1985, includes “ d ”, but refers to what we here call g):

$$\hat{\sigma}_2^2(g) = \frac{n_1 + n_2}{n_1 \cdot n_2} + \frac{g^2}{2 \cdot (n_1 + n_2)}. \quad (12a)$$

In terms of the general expression, it is:

$$\hat{\sigma}_2^2(g) = \frac{1}{\tilde{n}} + \frac{[c(m)]^2}{2 \cdot (n_1 + n_2)} \cdot d^2. \quad (13)$$

This estimator is implemented in many programs. For example, in *metafor* $\hat{\sigma}_2^2(g)$ can be selected as an estimator using the “vtype” argument and the “LS” option, but it is not necessary because this is the default option. Estimator $\hat{\sigma}_2^2(g)$ is recommended in many sources (e.g., Card, 2015; Fritz, Morris, & Richler, 2012; Grissom & Kim, 2012; Koricheva, Gurevitch, & Mengersen, 2013; Lipsey & Wilson, 2001; Petticrew, & Roberts, 2008; Pigott, 2012).

Direct substitution in the approximate formula of Hedges and Olkin (1985).

In *Review Manager* (2014), a resource widely used in health sciences, the way to estimate the variance of g is to take the approximate formula (8) from Hedges and Olkin (1985) and directly replace δ by g (Deeks & Higgins, 2010). The result is taken as an estimate of the variance of g , instead of d (which is how it was proposed in Hedges and Olkin, 1985). It is also the formula implemented in the R package *meta* (Schwarzer, Carpenter, & Rücker, 2015). In terms of our general expression, it would be:

$$\hat{\sigma}_3^2(g) = \frac{1}{\tilde{n}} + \frac{[c(m)]^2}{2 \cdot (n_1 + n_2 - 3.94)} \cdot d^2. \quad (14)$$

Substitution of d instead of g in the approximate formula and correction with $c(m)^2$.

In some calculation tools the estimation procedure proposed by Borenstein (2009; Borenstein et al, 2009; Borenstein & Hedges, 2019) is implemented. In this procedure formula (12a) is used, but by substituting d instead of g to obtain an estimate of the variance of d ; then (6) is used to obtain the estimate of the variance of g . As formula (12a) was not proposed as an estimator of the variance of d , but of g , to refer to this use we will designate it as formula (12b):

$$\hat{\sigma}^2(d) = \frac{n_1 + n_2}{n_1 \cdot n_2} + \frac{d^2}{2 \cdot (n_1 + n_2)}. \quad (12b)$$

We will see later that this procedure leads to systematic bias. Adjusting our general expression, (9), the estimator used with this procedure is:

$$\begin{aligned}\hat{\sigma}_4^2(g) &= [c(m)]^2 \cdot \left(\frac{1}{\tilde{n}} + \frac{d^2}{2 \cdot (n_1 + n_2)} \right) = \\ &= \frac{[c(m)]^2}{\tilde{n}} + \frac{[c(m)]^2}{2 \cdot (n_1 + n_2)} \cdot d^2\end{aligned}\tag{15}$$

This is the estimator that is recommended, for example, in *Comprehensive Meta-analysis*, although the program allows other options (Borenstein, Hedges, Higgins & Rothstein, 2013), or several *R* packages as, for example, *Compute Effect Size* (Del Re, 2015).

Comprehensive Meta-analysis allows one to directly enter the sample statistics of the primary studies as part of the meta-analysis, and to calculate both estimators, d and g , using (3) and (5). The two estimators are accompanied by estimates of their sampling variances (actually their standard errors). The program estimates the variance of d with the formula (12b), while that of g is obtained with the estimator $\hat{\sigma}_4^2(g)$.

Simple substitution of d instead of g in the approximate formula.

Due to the change in nomenclature in Hedges and Olkin (1985) we have already mentioned, some meta-analysts have confused the estimators and have substituted the biased estimator in formula (12a), which we call d here (Ellis, 2010). Perhaps they have considered that the variances estimated by substituting d or g are not sufficiently different to pay attention to such differences, or perhaps they have decided for another unknown reasons. Whatever the reason, our fifth estimator refers to the cases in which the point estimate is the unbiased one, since it is calculated with (5) and not with (3), but the variance is obtained directly with (12b) and is taken as the variance of g .

Adjusting our general expression, the estimation formula would be:

$$\hat{\sigma}_s^2(g) = \frac{1}{\tilde{n}} + \frac{1}{2 \cdot (n_1 + n_2)} \cdot d^2 \quad (16)$$

The *R* package *esc* (Lüdtke, 2019), for example, does the calculations in this way. It correctly returns the estimates d and g , as specified in the corresponding argument. However, the variance that it returns is the same for both: the one that provides (12b). A meta-analyst who uses this package and wishes to use the unbiased estimator would take the correct values of the estimator, g , but would associate them with variances calculated with the estimator $\hat{\sigma}_s^2(g)$.

The calculator developed by David Wilson³ is widely used in meta-analysis, often in connection with the macros developed by Lipsey and Wilson (2001). The option to calculate the *SMD* from means, variances and samples sizes provides the uncorrected estimator, d . The variance of d is calculated with (12b). It can be argued that this calculation only provides the descriptive value of d and that it is the user's responsibility to decide whether to correct d when the calculated value is to be used as estimator of the corresponding parameter. But even if g is calculated later, from $c(m)$ and the value of d provided by the calculator, if the variance provided by this calculator is used for the meta-analytic weighting, it is necessary to keep in mind that it is the variance estimated with (12b) and not with (12a). A user who does not pay attention to this fact will directly use the variances that this calculator returns and would then be using the estimator $\hat{\sigma}_s^2(g)$.

Table 1 summarizes the sample of computing resources that we have selected to illustrate the practical implementation of the five estimation methods.

³ <https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>

Table 1. Examples of computer programs that implement each of the estimators studied.

Estimator	Programs
$\hat{\sigma}_1^2(g)$ - Unbiased estimator of Hedges (1983)	R package <i>metafor</i> (vtype = “UB”)
$\hat{\sigma}_2^2(g)$ Approximate formula of Hedges and Olkin (1985)	R package <i>metafor</i> (vtype = “LS”)
$\hat{\sigma}_3^2(g)$ - Direct substitution in the approximate formula of Hedges and Olkin (1985)	<i>Review Manager</i> R package <i>meta</i>
$\hat{\sigma}_4^2(g)$ - Substitution of d instead of g in the approximate formula and correction with $c(m)^2$	<i>Comprehensive meta-analysis</i> R package <i>compute effect size</i>
$\hat{\sigma}_5^2(g)$ - Simple substitution of d instead of g in the approximate formula	R package <i>esc</i> David Wilson’s effect size calculator

Bias

In order to assess the bias, let's see what the expected value is of an estimator of the general form expressed in (9). Since for two given values of n_1 and n_2 the values of $K1$ and $K2$ are two constants, applying the properties of the expected value:

$$E[\hat{\sigma}^2(g)] = K1 + K2 \cdot E[d^2]. \quad (17)$$

To arrive at the expected value of any of the estimators that we are going to evaluate, we must obtain $E(d^2)$. Taking into account the expression (A4) of the appendix and substituting in (17) we have:

$$E[\hat{\sigma}^2(g)] = K1 + K2 \cdot \frac{m}{m-2} \left(\frac{1}{\tilde{n}} + \delta^2 \right). \quad (18)$$

To evaluate the bias of each estimator we have obtained its expected value by substituting in (18) its expressions for $K1$ and $K2$. These appear in Table 2.

Table 2. Expected values of the five estimators of the sampling variance assessed.

Estimator	$K1$	$K2$
$\hat{\sigma}_1^2(g)$	$1/\tilde{n}$	$\left(1 - \frac{1}{a}\right) \cdot [c(m)]^2$
$\hat{\sigma}_2^2(g)$	$1/\tilde{n}$	$[c(m)]^2 / 2(n_1 + n_2)$
$\hat{\sigma}_3^2(g)$	$1/\tilde{n}$	$[c(m)]^2 / 2(n_1 + n_2 - 3.94)$
$\hat{\sigma}_4^2(g)$	$[c(m)]^2 / \tilde{n}$	$[c(m)]^2 / 2(n_1 + n_2)$
$\hat{\sigma}_5^2(g)$	$1/\tilde{n}$	$1 / 2(n_1 + n_2)$

Figure 1 shows the bias of the estimators $\hat{\sigma}_2^2(g)$ to $\hat{\sigma}_5^2(g)$ as a function of the value of the parameter (δ) and the total sample size (N). It represents values of δ between 0 and 1.2, which is the most frequent range in psychology. To simplify the

figure, we have represented N as the total sample size and we have restricted ourselves to the case in which $n_1 = n_2 = N/2$. We have included values of N up to 60. With higher values of N the bias is greatly reduced with all five estimators. The bias is represented on the vertical axis. It is the difference between the expected value of the evaluated estimator and the parametric value according to (7). A simple algebraic manipulation (appendix) shows that the expected value of the estimator $\hat{\sigma}_1^2(g)$ equals the expression (7) and is therefore unbiased, as Hedges (1983) has already shown. That is why it does not appear in the figure. However, in order to be able to visually assess the bias, in each figure a plane is included at the value 0 of the vertical axis, which represents the lack of bias; that is, the value that would correspond to the estimator $\hat{\sigma}_1^2(g)$.

A first inspection shows that at least in the space studied, all but one of the studied estimators uniformly underestimate the sampling variance. The expected values of estimators, $\hat{\sigma}_2^2(g)$, $\hat{\sigma}_4^2(g)$, and $\hat{\sigma}_5^2(g)$ are always lower than the parameter, although there are areas where they are very close to it and where they can be considered free of bias in practice. The only one that shows a slight over-estimation in a small area of the studied space, but which at the same time is the one that shows the best general behavior is the estimator $\hat{\sigma}_3^2(g)$, implemented in programs like *Review Manager* or *meta*. The approximate formula, $\hat{\sigma}_2^2(g)$, which in many sources is recommended for large samples, shows inadequate behavior with small samples and it is somewhat arbitrary to establish a red line to define operationally what "large samples" are. Something similar happens with $\hat{\sigma}_5^2(g)$. The estimator $\hat{\sigma}_4^2(g)$ shows the highest levels of bias and in a greater range of the conditions studied. In fact, even with the largest sample sizes included here, the bias is not corrected.

Of course, in terms of bias the best option is $\hat{\sigma}_1^2(g)$. In *metafor* it is implemented in such a way that even the correction factor, $c(m)$, is not approximated according to formula (4), but is calculated with all the precision allowed by the expression reached by Hedges (1981; see footnote 1). In case of wanting to use a simpler method, the best alternative is $\hat{\sigma}_3^2(g)$, because its bias is very small and is limited to the smallest values of N and when δ is moderate or large.

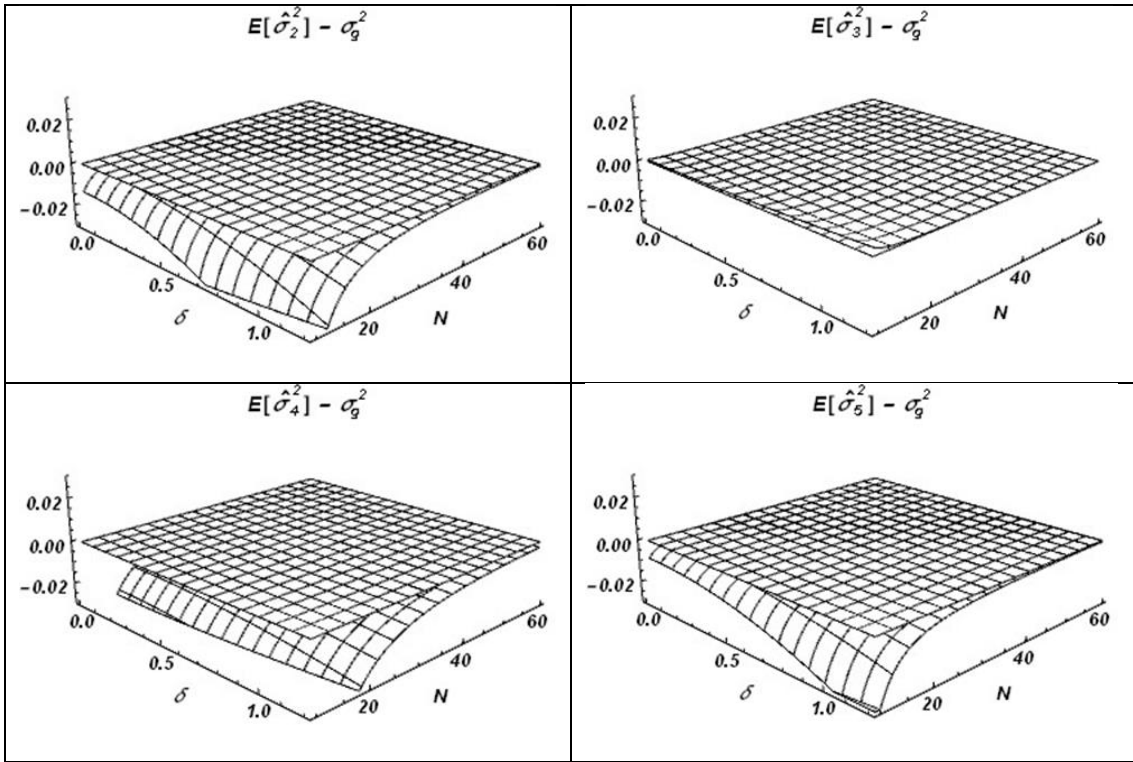


Figure 1. Bias of the estimators $\hat{\sigma}_2^2(g)$ to $\hat{\sigma}_5^2(g)$ as a function of the sample size and δ .

The plane at the value 0 represents an unbiased estimator, such as $\hat{\sigma}_1^2(g)$. The cuts in the edges reflect the fact that the expected value is only defined for $N > 4$ (see the appendix).

Variance

To evaluate and compare the estimators, we also have to pay attention to the variances. Given (9) and the expression (A7), which reflects the variance of d^2 (see the appendix):

$$\text{Var}[\hat{\sigma}_j^2(g)] = K2_j^2 \cdot \left[2 \cdot \frac{\left(\frac{1}{\tilde{n}} + \delta^2\right)^2 + \left(\frac{1}{\tilde{n}^2} + 2 \cdot \delta^2 \cdot \frac{1}{\tilde{n}}\right) \cdot (m-2)}{(m-2)^2 \cdot (m-4)} \cdot m^2 \right]. \quad (19)$$

As the only difference between the variances of the five estimators is due to the $K2$ factor, to compare their variances it is enough to compare these factors in Table 2. The comparison of those coefficients for the given values of δ and N allows us to order the estimators according to their variances:

$$\text{Var}(\hat{\sigma}_2^2) = \text{Var}(\hat{\sigma}_4^2) < \text{Var}(\hat{\sigma}_5^2) < \text{Var}(\hat{\sigma}_1^2) < \text{Var}(\hat{\sigma}_3^2)$$

They are represented in Figure 2. Of course, the surfaces do not intersect at any point, since they are ordered uniformly in efficiency, at least in the space of N and δ studied here. In terms of efficiency the estimators $\hat{\sigma}_2^2(g)$ and $\hat{\sigma}_4^2(g)$ are better than $\hat{\sigma}_1^2(g)$. The estimator $\hat{\sigma}_5^2(g)$ is somewhat worse, but also better than $\hat{\sigma}_1^2(g)$. The estimator $\hat{\sigma}_3^2(g)$, the best in terms of bias apart from $\hat{\sigma}_1^2(g)$, has a slightly higher variance (lower efficiency) with the smaller sample sizes (up to approximately $N=12$, that is, 6 per group).

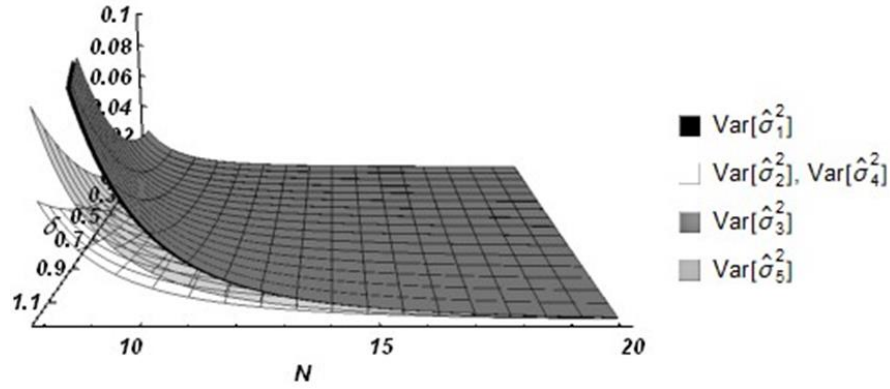


Figure 2. Variance of the estimators compared. The variances $\hat{\sigma}_2^2(g)$ and $\hat{\sigma}_4^2(g)$ are identical. The cuts in the edges reflect the fact that the variance is only defined for $N > 6$ (see the appendix).

Choosing an estimator

According to the properties of the estimators, the preferred estimator must be unbiased, consistent and of optimal efficiency (minimum variance). In our case, all the estimators are consistent, since a simple inspection of their coefficients $K1$ and $K2$ shows that these coefficients tend to 0 when N approaches infinity. The problem is that the best estimator in terms of bias, $\hat{\sigma}_1^2(g)$, is not the best estimator in terms of efficiency. If the estimator is chosen based on a combination of both the bias and the variance, such as the Mean Squared Error ($\text{MSE} = \text{variance} + \text{bias}^2$), the results are not uniform. For some combinations of δ and N the estimators are ordered in a certain way, while for other combinations that order changes. In the case of the five estimators of the sampling variance that we are comparing, we believe that the most important issue is the bias (see also White & Thomas, 2005). But whatever the option chosen by the meta-analysts, they must know that the practical consequences of the choice could not be trivial. As we will see in the next sections, the choice of an estimator can have

important consequences on the results of the meta-analysis, especially when fitting random-effects models.

Some examples based on real studies

It is difficult to assess the real impact of the differences between the levels of bias and the variances of the different estimators by simply observing the formulas in Table 2. In many cases the differences are very small, and are negligible for practical purposes. However, these small differences at the study level can lead to more important differences at the meta-analysis level. To make the impact of the choice between these procedures more visible, we have performed a sensitivity test with the meta-analyses included in the review by Rubio-Aparicio, Marin-Martinez, Sanchez-Meca and Lopez-Lopez (2018). This review includes a large number of meta-analyses on interventions in clinical psychology, of which 41 use *SMD* as the *ES* index. This set of meta-analyses has a number of studies ranging from 7 to 70, averaging $k = 24.2$. We have taken the primary studies from each of these meta-analyses and have obtained, by means of our own code (see the supplemental material), the five estimates of the sampling variance of each primary study using the formulas (10), (13), (14), (15) and (16). With them we have obtained combined *ES* estimates, first with a fixed effect model and then with a random effects model, in both cases using *metafor*.

Let's look at the results first with the fixed effect model⁴. The sampling variances estimated with the different procedures may be different, but if the inverses of the

⁴ Strictly speaking, a fixed effect model should not be fitted using the *g*-value of each study to estimate its sampling variance, but rather obtaining a single combined estimate (weighted by *N*) that is substituted in the calculations of the sampling variances of all studies (this is the *metafor* option `vtype = "AV"`). Here

variances with one method are proportional to those of another method, then the pooled estimate does not change. To evaluate the estimates obtained with each method of estimating the sampling variance, we have calculated the ratio between the combined estimate of methods 2 to 5 and that of method 1 (the unbiased one). We have also obtained the width of the confidence intervals and have divided the width calculated with each method with that of method 1. The results appear in Table 3.

Method 2 provides larger combined estimates than method 1, with excesses averaging 1.1% but reaching 7.3% in one meta-analysis. Method 3 presents combined estimates practically identical to those of method 1. Method 4 presents the largest mean deviation in the combined estimate, 1.8%, and it reaches up to 19.5% in one of the meta-analyses. Method 5 presents a slight mean overestimation (0.6%), but in one of the meta-analyses reaches 4.5%. The width of the confidence intervals is always very close to that of method 1, with the exception of method 4 which always shows slightly lower widths.

Table 3. Ratio between the combined estimates obtained with a fixed effect model with the methods 2 to 5 and that of method 1. The same results are shown for the width of the confidence intervals.

	Sampling variance method	Min	Max	Mean
g_{\bullet}	$\hat{\sigma}_2^2(g)$	1.001	1.073	1.011
	$\hat{\sigma}_3^2(g)$	0.998	1.000	1.000

we have used the options “UB” and “LS” for estimators 1 and 2 in order to facilitate comparisons with the results from the random effects model.

	$\hat{\sigma}_4^2(g)$	0.990	1.195	1.018
	$\hat{\sigma}_5^2(g)$	1.000	1.045	1.006
$UL_{g.} - LL_{g.}$	$\hat{\sigma}_2^2(g)$	0.991	1.000	0.998
	$\hat{\sigma}_3^2(g)$	1.000	1.000	1.000
	$\hat{\sigma}_4^2(g)$	0.961	0.996	0.984
	$\hat{\sigma}_5^2(g)$	0.995	1.000	0.999

In general, in psychology, the most frequent (and realistic) decision is to fit random effects models. The choice of a method for estimating the sampling variance will have a greater impact on the results than under a fixed effect model. Even if the inverses of the sampling variances obtained with two methods are proportional, the study weights are no longer proportional. The reason is that in a random effects model the weight is obtained by calculating the inverse of the sum of the sampling variance and the inter-study or specific variance, $\hat{w} = 1/[\hat{\sigma}^2(g) + \hat{\tau}^2]$. As the specific variance once estimated is a constant for calculations, the weights thus obtained with one method can no longer be proportional to those obtained with other methods, other than in the particular case in which the specific variance estimated is equal to 0 (fixed effect model).

To illustrate the impact of the estimation method with a random effects model, we have taken from the same base of meta-analyses those that have at least 20 studies, since with fewer studies the estimate of τ^2 is unstable (Langan et al, 2019). We have also eliminated two meta-analyses in which the estimate of the between-study variance is equal to 0. The specific variance has been estimated with the restricted maximum likelihood (*REML*) method. The results of the remaining 16 meta-analyses are shown in Table 4. The deviations of the combined estimates from those of method 1 are generally

greater than under a fixed effect model. The main exception is method 3, which again shows negligible deviations. The largest deviations are associated with method 4, with a mean deviation of 2.8%, which reaches 26.0% in one of the meta-analysis. Confidence intervals tend to show slightly larger amplitudes than with method 1, with the exception of method 3, which still shows width intervals very close to those of method 1. But the main difference is in the estimate of the specific variance. Method 3 returns estimates of τ^2 that are practically identical to those of method 1. That is why their differences in the rest of the meta-analyses are very small. With method 4, the specific variance is overestimated by 13.6% on average, but in one study it is 53.3%. Methods 2 and 5 also show estimates that exceed those of method 1 by 5.5% and 3.4% on average, but they reach 29.4% and 18.1%, respectively.

Table 4. Results with a subsample of 16 meta-analyses on interventions in clinical psychology. Ratios between the combined estimates obtained with a random effects model with methods 2 to 5 and method 1. The same results are reported for the widths of the confidence intervals and the estimates of the specific variance, τ^2 .

	Sampling variance method	Min	Max	Mean
$g.$	$\hat{\sigma}_2^2(g)$	1.000	1.101	1.014
	$\hat{\sigma}_3^2(g)$	0.997	1.000	1.000
	$\hat{\sigma}_4^2(g)$	0.998	1.260	1.028
	$\hat{\sigma}_5^2(g)$	1.000	1.062	1.008

$UL_{g.} - LL_{g.}$	$\hat{\sigma}_2^2(g)$	1.00	1.057	1.009
	$\hat{\sigma}_3^2(g)$	0.998	1.000	.9998
	$\hat{\sigma}_4^2(g)$	0.994	1.059	1.010
	$\hat{\sigma}_5^2(g)$	1.000	1.036	1.006
τ_j^2 / τ_1^2	$\hat{\sigma}_2^2(g)$	1.000	1.294	1.055
	$\hat{\sigma}_3^2(g)$	0.992	1.000	.9985
	$\hat{\sigma}_4^2(g)$	0.995	1.533	1.136
	$\hat{\sigma}_5^2(g)$	1.000	1.181	1.034

Discussion

The most popular tools for calculations in meta-analysis implement different methods to estimate the sampling variance of *SMD*. For a wide range of values of the parametric value, δ , and of sample sizes ($N = n_1 + n_2$) the differences between these methods at the study level are small. However, when studies are taken to obtain a pooled estimate of the *ES* with a random effects model, their impact can be quite large. The estimate of the specific variance, τ^2 , can vary greatly depending on the method used to estimate the sampling variance. This is because τ^2 is estimated essentially by subtracting the sampling variances from the sum of the squared deviations. If the sampling variances are underestimated, which is what happens in general with the estimators evaluated here (see Figure 1), then τ^2 is overestimated, and vice versa. In short, our main message is that when using *SMD* as the index of *ES*, *the results of a meta-analytic synthesis are sensitive to the decision about the method used to estimate the sampling variances of the ES estimates in the primary studies*. This is especially true

when fitting random effects models, which are the ones most widely used in psychology.

In terms of bias, there is no doubt in the choice: the unbiased estimator of Hedges (1983), $\hat{\sigma}_1^2(g)$, must be always preferred. Approximate formulas have been used for a long time, such as $\hat{\sigma}_2^2(g)$, which was developed at a time when meta-analyses were largely done using pocket calculators or meta-analysts had to program their own code. Then it was justified to lose precision in the estimates in exchange for greater simplicity and agility in the calculations. This saved time and reduced opportunities for mistakes. However, when a meta-analysis includes studies with large samples and studies with small samples, the easiest choice for a meta-analyst is to use a single formula and that formula is the simplest one. After all, there is a small step between "the formulas give very similar results if the samples are not too small" to "as this meta-analysis has some studies with large samples, it is justified to use the approximate formula because the difference will be small". Today there is no justification for using approximate formulas in general purpose programs (White & Thomas, 2005). In any case, if we want to use a simplified and approximate formula, we have a better alternative, the adaptation implemented in *Review Manager*, $\hat{\sigma}_3^2(g)$, which has a very small level of bias, negligible for practical purposes.

In terms of efficiency, we have found that there are several better estimators than $\hat{\sigma}_3^2(g)$, but we believe that in this case the level of bias should prevail as a priority criterion.

Although in most cases the differences between using one estimation method or another will be small, there will be some in which the difference will be larger. We have shown such effects on the basis of actual meta-analyses. This can be important for

different groups of people around the meta-analysis. First, the users of the programs do not have to be experts in statistical models for meta-analysis, so they will surely routinely use the default options of specific meta-analysis programs or the values provided by online calculators. Second, the consequences of choosing one of the methods to estimate the sampling variance can also be important to methodologists evaluating and comparing techniques, methods, and procedures through simulation studies. Choosing the default formulas or programming one or another formula can change the results and conclusions of the simulation studies. Especially when the study involves estimating the specific variance, τ^2 , under a random effects model. Finally, researchers or professionals who simply read the meta-analyzes as consumers of scientific literature should also be aware of the effects studied here. They should know that when certain approximate formulas have been used it is likely that the specific variance reported has been overestimated. Consequently, the confidence and credibility intervals could be distorted.

Some recommendations

We conclude this article with some practical recommendations in the development and design of calculation tools for meta-analysis. We have seen that the conclusions of a meta-analysis and the estimates it provides can be sensitive to the choice of an estimator of the sampling variance. We could establish some rules for choosing an estimator, based on the sample sizes and the combined *ES* estimate. However, there are three reasons why we think it is better to set simple and, if possible, unique rules. First, setting threshold values of the sample sizes and the combined estimate of the *ES* to choose one or the other estimator is always somewhat arbitrary. Second, the main alleged criterion, which is essentially the sample size, can vary greatly between studies

within a meta-analysis. Obviously, it does not seem reasonable to use different methods for the primary studies within the same meta-analysis. Sampling variance should be estimated with the same method for all studies in a given meta-analysis, which would sometimes lead to ambiguous situations if it includes studies with small and large samples. Third, many meta-analysts typically use different programs to apply the different techniques they use in the same meta-analysis (e.g., for pooled estimation, moderator analysis, figures, or publication bias). For the results to be congruent within a given meta-analysis, it is necessary for the different programs to use the same method of estimating the sampling variance.

On the other hand, we have seen in the set of re-analyzed meta-analyses that the results of some of them change according to the chosen estimator, but not others. Since we do not know when in a meta-analysis the differences are going to be important, the best decision is not to risk that they are. It is better to implement methods that are always correct, since nowadays they do not suppose a relevant increase in computing time. Therefore, these are our recommendations:

1. Implement the unbiased formula, estimator $\hat{\sigma}_1^2(g)$. In programs that already have this method implemented, such as *metafor*, we suggest considering that it also be programmed as the default. This reduces the probability that the approximate formula will be abused.
2. Although we do not conceive of any situation in which a method other than unbiased might be preferable, perhaps a researcher may decide to use another method for their own reasons. For example, because you are writing your own code and you don't have much programming skills. In these situations, the estimator $\hat{\sigma}_3^2(g)$, which is implemented in programs such as *review manager* or *meta*, should be chosen as the first option. The estimator $\hat{\sigma}_2^2(g)$, which appears in most texts on

meta-analysis, should be replaced by $\hat{\sigma}_3^2(g)$, since the level of bias of this is clearly lower.

3. The method designated here as the estimator $\hat{\sigma}_4^2(g)$ (implemented, for example, in *comprehensive meta-analysis*) should be modified, as it offers estimates that systematically underestimate the sampling variance and can have important consequences when fitting random-effects models.
4. The procedure implemented as the estimator $\hat{\sigma}_5^2(g)$ should be completely avoided, as it also shows a high bias. Furthermore, it is incongruous to use a point estimator, g , and then associate to it the variance of another estimator, d .

References

- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (eds.), *The Handbook of Research Synthesis*, 2^a ed. (pp. 221-235). New York: Russell Sage Foundation.
- Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-Analysis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-analysis* (3rd ed., pp. 383–429). New York: Russell Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-analysis*. Chichester, UK: John Wiley and Sons.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effects and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97-111.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). *Comprehensive Meta-analysis Version 3.0*. Biostat Inc.

- Card, N. A. (2015). *Applied Meta-analysis for Social Science Research*. Guilford Publications.
- Deeks, J. J., & Higgins, J. P. (2010). Statistical algorithms in review manager 5. *Statistical Methods Group of The Cochrane Collaboration*, 1(11).
- Del Re, A. C. (2015). *Compute effect sizes (Version 0.2-4)* [Software]. Retrieved from <http://cran.r-project.org/web/packages/compute.es/compute.es.pdf>
- Ellis, P. D. (2010). *The Essential Guide to Effect Sizes*. New York: Cambridge University Press.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.
- Grissom, R. J., & Kim, J. J. (2012). *Effect Sizes for Research: Univariate and Multivariate Applications, 2nd ed.* Lawrence Erlbaum Associates Publishers.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
- Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93(2), 388-395.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press.
- Johnson, N., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions (2nd Ed., Vol. 2)*. New York: Wiley.
- Koricheva, J., Gurevitch, J., & Mengersen, K. (Eds.). (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.
- Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., ... & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in

- simulated random-effects meta-analyses. *Research Synthesis Methods*, 10(1), 83-98.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis*. Thousand Oaks, CA: Sage.
- Lüdtke, D. (2019). *Esc: Effect Size Computation for Meta-Analysis* (Version 0.5.1).
<https://doi.org/10.5281/zenodo.1249218>
- Petticrew, M., & Roberts, H. (2008). *Systematic Reviews in the Social Sciences: A Practical Guide*. John Wiley & Sons.
- Pigott, T. (2012). *Advances in meta-analysis*. Heidelberg, Germany: Springer Science & Business Media.
- Review Manager (RevMan) (2014) [Computer program]. *Version 5.3*. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.
- Rubio-Aparicio, M., Marín-Martínez, F., Sánchez-Meca, J., & López-López, J. A. (2018). A methodological review of meta-analyses of the effectiveness of clinical psychology treatments. *Behavior Research Methods*, 50(5), 2057-2073.
- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). New York: Springer.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. doi:10.18637/jss.v036.i03
- White, I. R., & Thomas, J. (2005). Standardized mean differences in individually-randomized and cluster-randomized trials, with applications to meta-analysis. *Clinical Trials*, 2(2), 141-151.

APPENDIX

Variance of d

Knowing that the formula for the parametric variance of g is (7) and substituting g according to (5), the variance of d is:

$$\sigma^2(d) = \sigma^2\left(\frac{g}{c(m)}\right) = \frac{1}{[c(m)]^2} \left[\frac{a}{\tilde{n}} \cdot (1 + \tilde{n} \cdot \delta^2) - \delta^2 \right].$$

Substituting a according to its definition, $[c(m)]^2 \cdot m/(m-2)$, the expression that appears in Hedges (1981) and Hedges and Olkin (1985) is reached:

$$\sigma^2(d) = \frac{m}{(m-2) \cdot \tilde{n}} + \delta^2 \cdot \left[\frac{m}{m-2} - \frac{1}{[c(m)]^2} \right].$$

Expected value and variance of d^2 .

Following Hedges (1981), the effect size estimator d can be expressed as [recall that $\tilde{n} = (n_1 n_2)/(n_1 + n_2)$]:

$$d = w \cdot \frac{1}{\sqrt{\tilde{n}}}, \tag{A1}$$

where w is a continuous random variable that follows a non-central Student's t distribution with m degrees of freedom (df) and non-centrality parameter $\delta \cdot \sqrt{\tilde{n}}$.

Squaring at (A1):

$$d^2 = w^2 \cdot \frac{1}{\tilde{n}}. \quad (\text{A2})$$

If $y = w^2$, then the variable y is a continuous random variable that follows a non-central F distribution with $df1 = 1$ and $df2 = m$, and non-centrality parameter $\delta^2 \cdot \tilde{n}$ (Johnson, Kotz, & Balakrishnan, 1994). From here:

$$E[d^2] = E[y] \cdot \frac{1}{\tilde{n}}. \quad (\text{A3})$$

Now, it is known that the expected value of a continuous random variable that follows a non-central F distribution with $df1$ and $df2$ (where $df2 > 2$) and non-centrality parameter λ is: $\frac{df2 \cdot (df1 + \lambda)}{df1 \cdot (df2 - 2)}$ (Johnson, Kotz, & Balakrishnan, 1994). In the present situation $df1 = 1$, $df2 = m$, and $\lambda = \delta^2 \cdot \tilde{n}$. Then from (A3) follows:

$$E[d^2] = \frac{m}{m-2} \cdot (1 + \delta^2 \cdot \tilde{n}) \cdot \frac{1}{\tilde{n}} = \frac{m}{m-2} \cdot \left(\frac{1}{\tilde{n}} + \delta^2 \right). \quad (\text{A4})$$

With respect to the variance of the estimator, from (A2) and the definition of y it follows that:

$$Var[d^2] = \left[\frac{1}{\tilde{n}} \right]^2 \cdot Var[y]. \quad (\text{A5})$$

The variance of a random variable following an F distribution (being $df2 > 4$) is (Johnson, Kotz, & Balakrishnan, 1994):

$$2 \cdot \frac{(df1 + \lambda)^2 + (df1 + 2 \cdot \lambda)(df2 - 2)}{(df2 - 2)^2 \cdot (df2 - 4)} \cdot \left(\frac{df2}{df1} \right)^2. \quad (A6)$$

Substituting in (A5) and simplifying:

$$Var[d^2] = 2 \cdot \frac{\left(\frac{1}{\tilde{n}} + \delta^2 \right)^2 + \left(\frac{1}{\tilde{n}^2} + 2 \cdot \delta^2 \cdot \frac{1}{\tilde{n}} \right) \cdot (m - 2)}{(m - 2)^2 \cdot (m - 4)} \cdot m^2. \quad (A7)$$

Expected value of the estimator $\hat{\sigma}_1^2(g)$ of Hedges (1983)

As we already know that the parametric variance of g is that expressed in (7), to show that it is an unbiased estimator it is enough to show that its expected value is equal to that expression. Substituting the expected value of d^2 in (11):

$$E(\hat{\sigma}_1^2) = \frac{1}{\tilde{n}} + \left(1 - \frac{1}{a} \right) \cdot [c(m)]^2 \cdot \frac{m}{m - 2} \cdot \left(\frac{1}{\tilde{n}} + \delta^2 \right).$$

With some basic algebra we arrive at (7), which is the formula derived by Hedges (1983):

$$\begin{aligned}
E\left(\hat{\sigma}_1^2\right) &= \frac{1}{\tilde{n}} + \left(\frac{a-1}{a}\right) \cdot a \cdot \left(\frac{1}{\tilde{n}} + \delta^2\right) = \\
&= \frac{a}{\tilde{n}} + (a-1) \cdot \delta^2 = \\
&= \frac{a}{\tilde{n}} \cdot (1 + \tilde{n} \cdot \delta^2) - \delta^2
\end{aligned}$$