

Negative estimate of variance-accounted-for effect size: How often it is obtained, and what happens if it is treated as zero

Kensuke Okada¹ 

Published online: 22 June 2016
© Psychonomic Society, Inc. 2016

Abstract Researchers recommend reporting of bias-corrected variance-accounted-for effect size estimates such as omega squared instead of uncorrected estimates, because the latter are known for their tendency toward overestimation, whereas the former mostly correct this bias. However, this argument may miss an important fact: A bias-corrected estimate can take a negative value, and of course, a negative variance ratio does not make sense. Therefore, it has been a common practice to report an obtained negative estimate as zero. This article presents an argument against this practice, based on a simulation study investigating how often negative estimates are obtained and what are the consequences of treating them as zero. The results indicate that negative estimates are obtained more often than researchers might have thought. In fact, they occur more than half the time under some reasonable conditions. Moreover, treating the obtained negative estimates as zero causes substantial overestimation of even bias-corrected estimators when the sample size and population effect are not large, which is often the case in psychology. Therefore, the recommendation is that researchers report obtained negative estimates as is, instead of reporting them as zero, to avoid the inflation of effect sizes in research syntheses, even though zero can be considered the most plausible value when interpreting such a result. R code to reproduce all of the described results is included as [supplemental material](#).

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0760-y) contains supplementary material, which is available to authorized users.

✉ Kensuke Okada
ken@psy.senshu-u.ac.jp

¹ Department of Psychology, Senshu University, 2-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa, Japan 214-8580

Keywords Effect size · Variance-accounted-for · Eta squared · ANOVA

The reporting of effect size estimates has been advocated by many psychology journal editors and authors, as well as by the *APA Publication Manual* (Fritz, Morris, & Richler, 2012; Huberty, 2002; Kline, 2013; Lakens, 2013). Thus, psychologists have recently adopted a common practice of reporting variance-accounted-for effect size estimates together with analysis of variance (ANOVA) results. The *population* variance-accounted-for effect size is defined as the proportion of the total population variance (of the dependent variable) that is explained by the factor of interest. Since the population variance is generally unknown, we need to estimate the population effect on the basis of sample data. Eta squared (η^2 ; Fisher, 1925), omega squared (ω^2 ; Hays, 1963), and epsilon squared (ϵ^2 ; Kelley, 1935) are three major *sample* effect size estimators¹ used in psychology, as well as in other sciences. They are summarized in Table 1.

One important distinction among them, shown in Table 1, is whether or not they are bias-corrected. The bias of an estimator is defined in statistics as the difference between the mean of the sampling distribution of the estimator, which is also called the *expected value*, and the true population value. The η^2 statistic is an *uncorrected* effect size estimator, which is simply defined as the proportion of the total sum of squares accounted for by the factor. Although its simplicity is an advantage, η^2 is known to have an upward bias, meaning that on average it overestimates the population effect, especially when the sample size is small (Levine & Hullett, 2002).

¹ Note that the terms “estimate” and “estimator” are different. An *estimate* is a number that is the best guess for the population value, and an *estimator* is a function for calculating an estimate from given data.

Table 1 Formulas for three sample variance-accounted-for effect size estimators in one-factor between-subjects designs

Sample Effect Size	Bias Correction	Formula	Reference
Eta squared (η^2)	Uncorrected	$\eta^2 = \frac{SS_M}{SS_T}$	Fisher (1925)
Omega squared (ω^2)	Corrected	$\omega^2 = \frac{SS_M - df_M MS_E}{SS_T + MS_E}$	Hays (1963)
Epsilon squared (ε^2)	Corrected	$\varepsilon^2 = \frac{SS_M - df_M MS_E}{SS_T}$	Kelley (1935)

SS sum of squares, *df* degrees of freedom, and *MS* mean squares. For the subscripts, *T* total, *M* treatment (or means), and *E* error

On the other hand, ω^2 and ε^2 are given by correcting the bias in estimating the population effect. Statistically, no *unbiased* estimator (meaning an estimator with exactly zero bias) of the population variance-accounted-for effect size is known. However, ω^2 and ε^2 are each derived to make the bias as small as possible (Winkler & Hays, 1975). Therefore, they are called *bias-corrected* effect size estimators. These two bias-corrected estimators share the same basic idea—replacement of unknown quantities in the population with their unbiased sample estimators—although their final forms differ in their denominators (see Table 1). As is described in detail by Glass and Hakstian (1969), this difference stems from the fact that ω^2 and ε^2 relate to different decompositions of the population effect size. Put simply, Kelley (1935) derived ε^2 by simply rewriting the population formula using the total and error variance in the population, and then replacing them with their unbiased sample estimators. On the other hand, Hays (1963) first rewrote the formula of the population effect size by explicitly considering the assumptions of the ANOVA model, and then replaced both the denominator and the numerator with their respective unbiased estimators. Thus, it is not that one is right and the other wrong; rather, both of the bias-corrected estimators have different grounds. Readers who are interested in complete derivations and comparative assessment of the formulas may refer to Glass and Hakstian (1969).

Because of the nonnegligible positive bias of η^2 , use of the bias-corrected estimators ω^2 and ε^2 has been recommended by statistical textbooks and statistically minded researchers (Ferguson, 2009; Fritz et al., 2012; Grissom & Kim, 2012; Keppel & Wickens, 2004; Maxwell & Delaney, 2004). Previous simulation study results (Carroll & Nordholm, 1975; Keselman, 1975; Okada, 2013; Skidmore & Thompson, 2013) appear to support these recommendations. For example, Skidmore and Thompson conducted a simulation study to evaluate the bias of the effect size estimators when assumptions of ANOVA are violated, and summarized their results as follows: “Overall, our results corroborate the limited previous research (Carroll & Nordholm, 1975; Keselman, 1975) and suggest that η^2 should not be used as an ANOVA effect size estimator, because across the range of conditions we examined, η^2 had considerable sampling error

bias” (p. 544). On the other hand, the bias of ω^2 and ε^2 is generally known to be small.

However, these former studies miss an important fact. That is, bias-corrected effect size estimators, both ω^2 and ε^2 , can take negative values. In other words, the sampling distributions of ω^2 and ε^2 include ranges below zero. This issue is a side effect of bias correction; the uncorrected estimator, η^2 , never takes a value below zero. Of course, a negative variance ratio does not make sense in the real world because the ratio of explained to total variance must lie between 0 and 1. Therefore, when a negative estimate is obtained in practice, it is typically reported as zero (Kenny, Kashy, & Cook, 2006; Olejnik & Algina, 2000). Some textbooks also recommend this practice; for example, Keppel and Wickens (2004) state that when ω^2 gives a negative value, “it is best to set $\omega^2 = 0$ ” (p. 164).

Thus, theory and practice have been separated until this point. In theory, the bias of the corrected effect size estimator is small, given that both positive and negative estimates are reported exactly as defined in Table 1; in practice, a negative estimate is treated and reported as zero. The existing literature on variance-accounted-for effect sizes does not appear to have taken the issue of negative effect size estimates very seriously. However, after negative estimates had been obtained several times in empirical data analysis in the present lab, this issue seemed worth investigating. The goal of this study is to investigate in general (1) how often negative variance-accounted-for effect size estimates are obtained in a simple ANOVA model under reasonable conditions, and (2) how biased the estimators are under the typical reporting practice, that is, when negative effect size estimates are reported as zero. A simulation study is conducted to find the answers.

Method

Following previous simulation studies (Keselman, 1975; Okada, 2013), we consider a one-factor, between-subjects ANOVA design with four levels and manipulate three experimental factors: (a) three levels of population effect size, (b) two levels of population mean variability, and (c)

three levels of sample size. Each manipulated factor is described below.

Population effect size

In interpreting effect size, Cohen (1988) provided a guideline about small, medium, and large effect size values based on a number of study results in behavioral sciences. Of course, the meaning of the same effect size value would differ depending on the field and context, and many authors, including Cohen himself, have argued that the guideline needs to be treated with caution (Schagen & Elliot, 2004). However, Cohen's guideline is irrefutably well-known and has been adopted by a number of authors (Aguinis, Beaty, Boik, & Pierce, 2005). Therefore, here three conditions of population variance-accounted-for effect size values are investigated, corresponding to Cohen's small, medium, and large criteria: $\eta_{\text{pop}}^2 = .010, .059, \text{ and } .138$, respectively.

Population mean variability

Cohen (1988) also provides patterns of population means exhibiting different degrees of variability, such as *maximum variability*, in which the means are dichotomized at the end points of their range, and *intermediate variability*, in which the means are spaced equally over the range. Following previous simulation studies (Keselman, 1975; Okada, 2013), two conditions were included as experimental factors. Although Cohen also considered minimum variability, it is not used in the present study. It is relevant only with a larger number of levels than in this study. The previous studies shown above also used only the maximum and intermediate variability conditions.

Sample size

Marszalek, Barber, Kohlhart, and Holmes (2011) exhaustively investigated the sample size used in published experimental psychological research. According to their most recent survey, the 25 %, 50 %, and 75 % quantiles of the total sample size in published research are, respectively, 10, 18, and 32, and the sample sizes per group are 10, 12, and 19. Considering this, three sample sizes were investigated, $N = 20, 40, \text{ and } 80$, corresponding to the per-group sample sizes of 5, 10, and 20, respectively.

Thus, $3 \times 2 \times 3 = 18$ conditions in total were created. The actual population means that correspond to the 3×2 population effect size and population mean variability conditions are summarized in Table 2, with another three sample size conditions for each cell. For each condition, a random sample of the specified size was repeatedly generated from the normal population, with a mean as shown in Table 2 and variance one. Then, three sample effect size estimates were computed and

stored. This process was replicated 1,000,000 times per condition. All R codes used in this study to conduct the simulation and display the results are provided as [supplemental material](#).

Results

Ratio of negative estimates

The resultant ratios of negative bias-corrected effect size estimates, for both ω^2 and ε^2 , are summarized in Table 3 for all conditions. By definition, the signs of ω^2 and ε^2 are always the same.² Therefore, the ratios of negative estimates are exactly the same for both of them.

Rather surprisingly, more than half of the estimates produce negative values when the population effect size is small and the total sample size N is 20 or 40, in both population mean variability conditions. Likewise, more than 40 % of the estimates result in negative values when the population effect is small and $N = 80$, as well as when the population effect is medium and $N = 20$. The percentage of negative estimates decreases as the population effect size and sample size increase. However, the results in Table 3 generally suggest that negative effect size estimates are obtained more often than most psychology researchers might have thought.

Because population mean variability does not substantially affect the resultant effect size estimate, hereafter, only the results under the intermediate mean variability condition are described. The plots under the maximum mean variability condition are shown in the [supplemental material](#).

Figure 1 presents histograms of the three sample effect size estimates, to show how their distributions differ. The shaded bars indicate negative estimates. The distributions of the two bias-corrected estimators (ω^2 and ε^2) are very similar. Comparing them with the uncorrected η^2 , you can see that the overall shapes of the distributions are similar within the same condition. However, they obviously differ in their x -axis positions. The bias-corrected estimates have substantial mass in the negative region, especially when the population effect and the sample size are small, whereas η^2 has, by definition, no mass in the negative region. Note that this does not necessarily mean that η^2 is a better estimator, as will be described shortly.

Bias when negative estimates are treated as zero

Figure 2 shows boxplots of the effect size estimators for a bias comparison when negative estimates are treated as zero and

² From the definitions of ω^2 and ε^2 in Table 1, the numerators are, clearly, common to both estimates, and the difference lies in the denominators. Considering that the sum of squares in the denominators is always above zero, the sign of the estimates is determined by the numerator, which is common to both. Therefore, the signs of the two estimates always match.

Table 2 Settings of the population means ($\mu_1, \mu_2, \mu_3,$ and μ_4) in three population effect size and two population mean variability conditions

Population Effect Size		Population Mean Variability							
Interpretation	η^2_{pop}	Intermediate				Maximum			
		μ_1	μ_2	μ_3	μ_4	μ_1	μ_2	μ_3	μ_4
Small	$\eta^2_{pop}=.010$.000	.089	.179	.268	.000	.000	.200	.200
Medium	$\eta^2_{pop}=.059$.000	.224	.447	.671	.000	.000	.500	.500
Large	$\eta^2_{pop}=.138$.000	.358	.716	1.073	.000	.000	.800	.800

The values shown above are rounded to the nearest three decimal places. Originally, Cohen (1988) defined his guideline in the metric of f , where $\eta^2_{pop} = \frac{f^2}{1+f^2}$. Thus, in the present simulation, his original f criteria ($f = .10, .25,$ and $.40$ as small, medium, and large effect) were used and transformed to η^2_{pop} , to avoid the effect of rounding errors

when they are not. Here, two plots per simulation condition are shown for the bias-corrected estimators. Those with asterisks ($\omega^{2*}, \epsilon^{2*}$) indicate results when negative estimates are treated as zero, whereas those without asterisks (ω^2, ϵ^2) indicate results when all the obtained estimates, including the negative estimates, are treated just as they are. The results of the uncorrected estimator (η^2) are also shown.

When comparing the means of the sampling distributions (“x” mark) and the population values (vertical thick dashed line), we note two findings consistent with previous studies (Keselman, 1975; Okada, 2013). First, η^2 substantially overestimates the population effects in all conditions, and second, the bias of the corrected estimators (ω^2 and ϵ^2) is small, if not zero.

Moreover, the new findings are based on negative estimates treated as zero. In this case, the means of the corrected estimators (ω^{2*} and ϵ^{2*}) are systematically larger than the population effect when the population effect and the sample size are not large. This means that the practice of treating a negative estimate as zero causes considerable overestimation. The degree of overestimation increases as the population effect size and sample size decrease, in accordance with the ratios of negative estimates. Even though the biases of ω^{2*} and ϵ^{2*} are smaller than that of η^2 , bias is still substantial when the population effect size and sample size are small.

For example, when the population effect size is small—that is, when $\eta^2_{pop} = .010$ —the sample means of ω^{2*} and ϵ^{2*} are .057 and .059 when $N = 20$. This means that they provide estimates almost six times as large, on average. The degree of overestimation decreases as the sample size increases, but the estimates are three times as large when $N = 40$, and 1.9 times as large when $N = 80$. Thus, the practice of reporting negative effect size estimates as zero clearly results in substantial inflation of the estimators, especially when the population effect and sample size are small. On the other hand, when both the population effect and sample size are large, the corrected estimators’ bias is small, irrespective of the reporting practice, because in this case the estimate rarely takes a negative value.

Discussion

This article is based on a simulation study investigating how often negative effect size estimates are encountered and what is the consequence when they are treated as zero, as in current reporting practice. The findings can be summarized as follows.

First, negative variance-accounted-for effect size estimates are naturally obtained more often than most researchers would have thought. In fact, more than half of the estimates are

Table 3 Ratios of negative bias-corrected effect size estimates obtained from the simulation study

Population Effect Size		Population Mean Variability					
Interpretation	η^2_{pop}	Intermediate			Maximum		
		$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$
Small	$\eta^2_{pop} = .010$.554	.539	.491	.555	.539	.491
Medium	$\eta^2_{pop} = .059$.426	.309	.151	.426	.310	.152
Large	$\eta^2_{pop} = .138$.255	.102	.013	.256	.101	.013

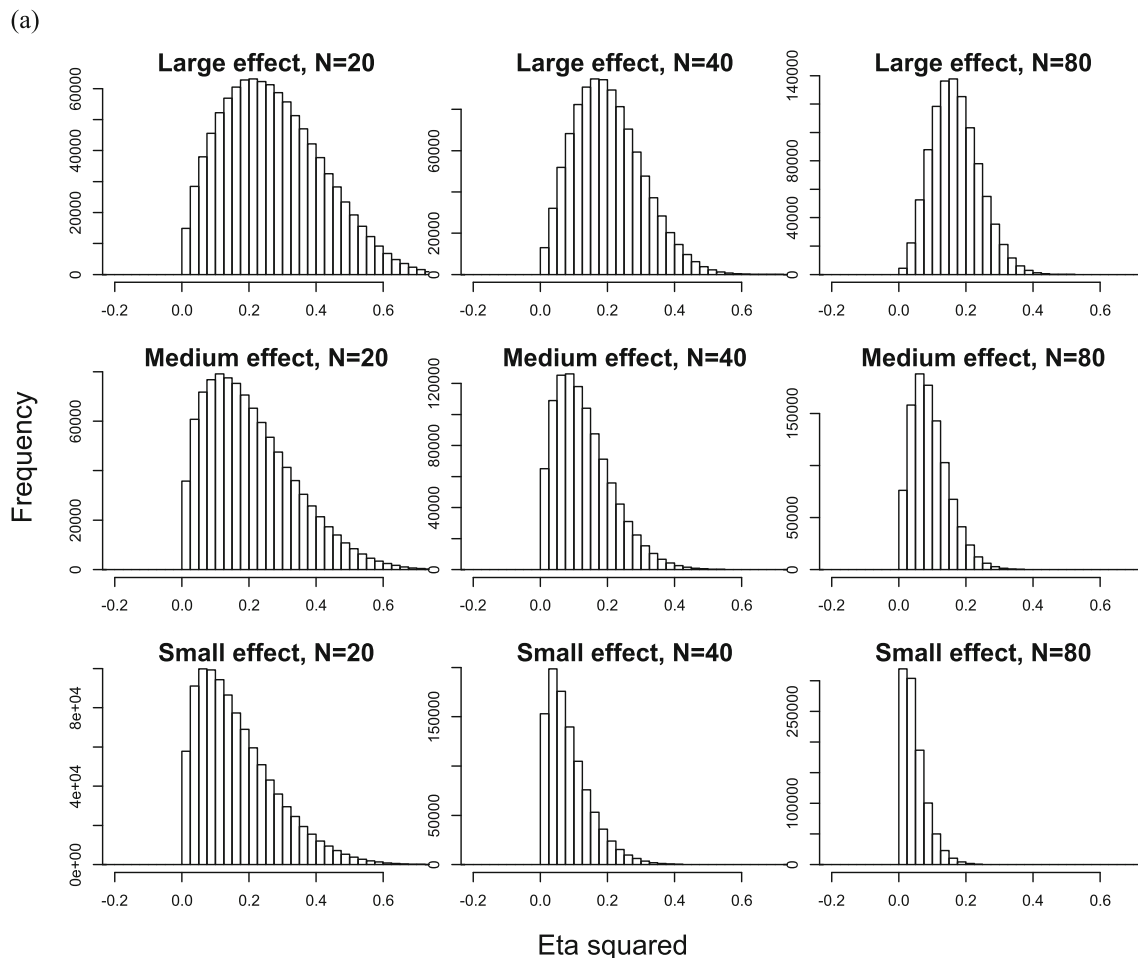


Fig. 1 Resultant histograms of (a) η^2 , (b) ω^2 , and (c) ε^2 in the intermediate population mean variability condition. The shaded bars indicate that the area is below zero

negative under some reasonable conditions. They are more likely to be obtained when the population effect and sample size are small. Second, treating the obtained negative effect size estimate as zero causes overestimation of even bias-corrected estimators. The degree of overestimation is substantial when the population effect and sample size are not very large, consistent with the first finding.

Therefore, in reporting the sample bias-corrected estimate, researchers should not hesitate to report the obtained negative effect size estimates. Negative estimates are not due to an error in analysis or data coding, but are naturally obtained just because of sampling fluctuations. If researchers do not report that the obtained estimate is negative, but report it as zero, readers would have difficulty in determining whether the obtained estimate is actually zero or a negative value reported as zero. Readers might be able to reconstruct the actual estimated value from other reported statistics, but this possibility can hardly be taken for granted. Treating a negative estimate as zero may also lead to overestimation of the effect in research synthesis studies such as meta-analysis because of estimator bias in each individual study. Considering that research

synthesis studies are generally believed to be more accurate and reliable, this could lead to an undesirable consequence.

Hence, even though a negative variance ratio does not make sense, it is recommended that researchers report the obtained negative estimate as it is. Reporting the obtained negative estimate as zero is not recommended, because this practice leads to a biased estimator and the actually obtained estimate would then be lost. This recommendation applies not only to point estimates but also, for the same reason, to interval estimates such as confidence intervals. Several methods exist to construct confidence intervals for the bias-corrected estimator (Finch & French, 2012), but their confidence limits could be negative, regardless of the specific method chosen, because a sampling distribution of effect sizes often includes regions below zero, as has been shown in this article.

Note that this article recommends *reporting* negative estimates obtained to reduce the overall bias in estimating the effect size, especially from a research synthesis perspective and a meta-analytic approach. On the other hand, the *interpretation* of the obtained negative estimate is conceptually challenging, because in reality the ratio of

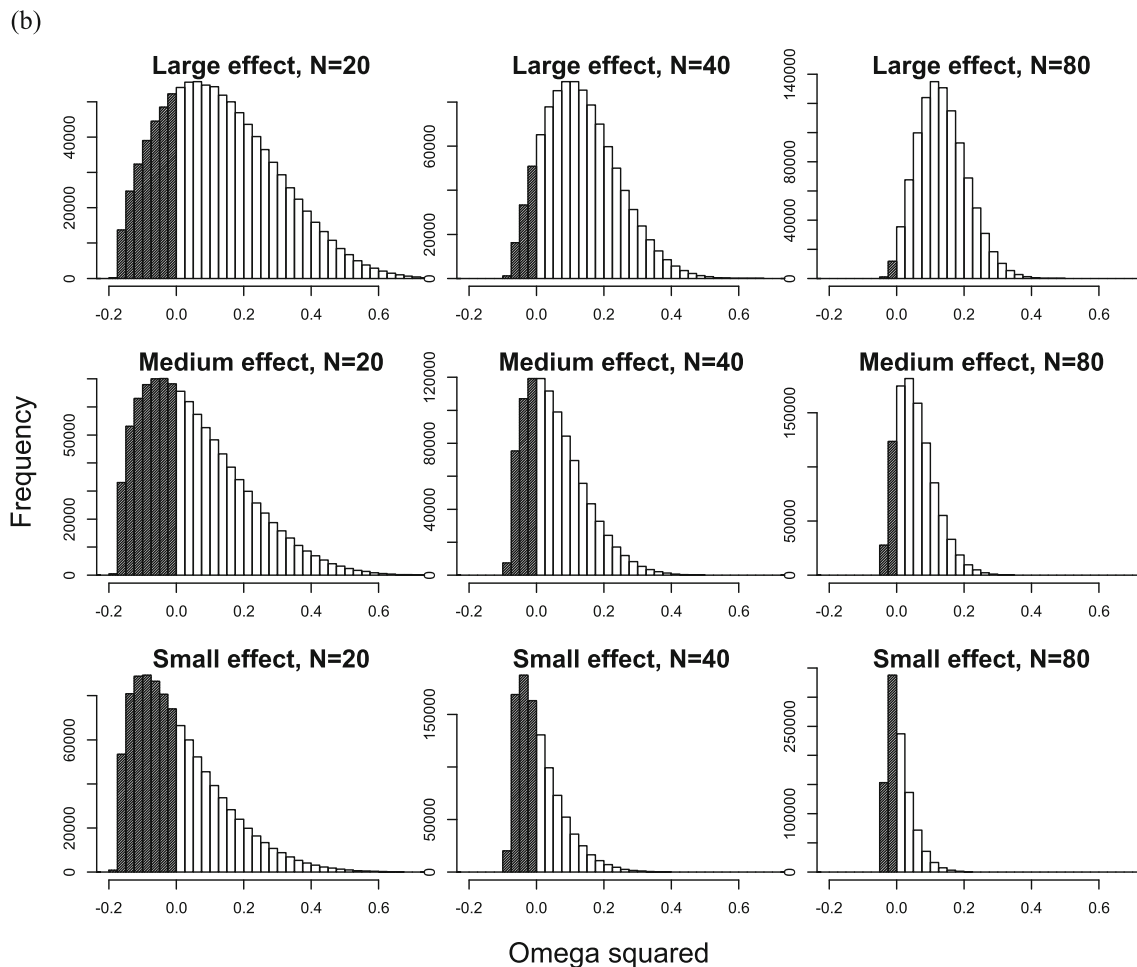


Fig. 1 (continued)

explained to total variance should lie between 0 and 1. Here are some suggestive interpretations. From the results in this article, negative estimates are likely to result under the following two scenarios. When a negative estimate is obtained for a small sample size, it is likely to be due to random sampling fluctuations, and different estimates may be obtained in the replication study. Thus, in this case, it may be better to increase the sample size in order to obtain a more reliable estimate. On the other hand, when a negative estimate is obtained for a large sample, it could mean that the true effect is close to zero.

In addition, negative bias-corrected estimators do not necessarily imply that using uncorrected η^2 is better. Even though η^2 , by definition, does not take negative values, it substantially overestimates the population effect, especially when the sample size and population effect are small. Fritz et al. (2012) argue that researchers tend to report uncorrected rather than corrected effect size estimates because the former are usually much larger than the latter, so that the results “look good.” However, this is because of its upward bias, so the reporting of an uncorrected estimate is not recommended either.

The take-home message of this study is simply stated as follows: do not hesitate to report a negative bias-corrected effect size estimate obtained as it is. In fact, this point has already been made by Lakens (2015). He states that “although effect sizes in the population that represent the proportion of variance are bounded between 0 and 1, unbiased effect size estimates such as ω^2 can be negative. Some people are tempted to set ω^2 to 0 when it is smaller than 0, but for future meta-analyses it is probably better to just report the negative value, even though it is impossible.” The contribution of the present article may still be significant, because it has quantitatively shown that negative estimates are easily obtained in practice and can result in substantially biased estimator when they are treated as zero. Because the effect size is sometimes considered the most important outcome of empirical studies (Lakens, 2013), the findings and recommendations above will be relevant to the reporting of various types of psychological study results.

The sample size considered in this study may not look very large. Actually, the issue of negative estimates would not be a problem with large samples, because the bias-

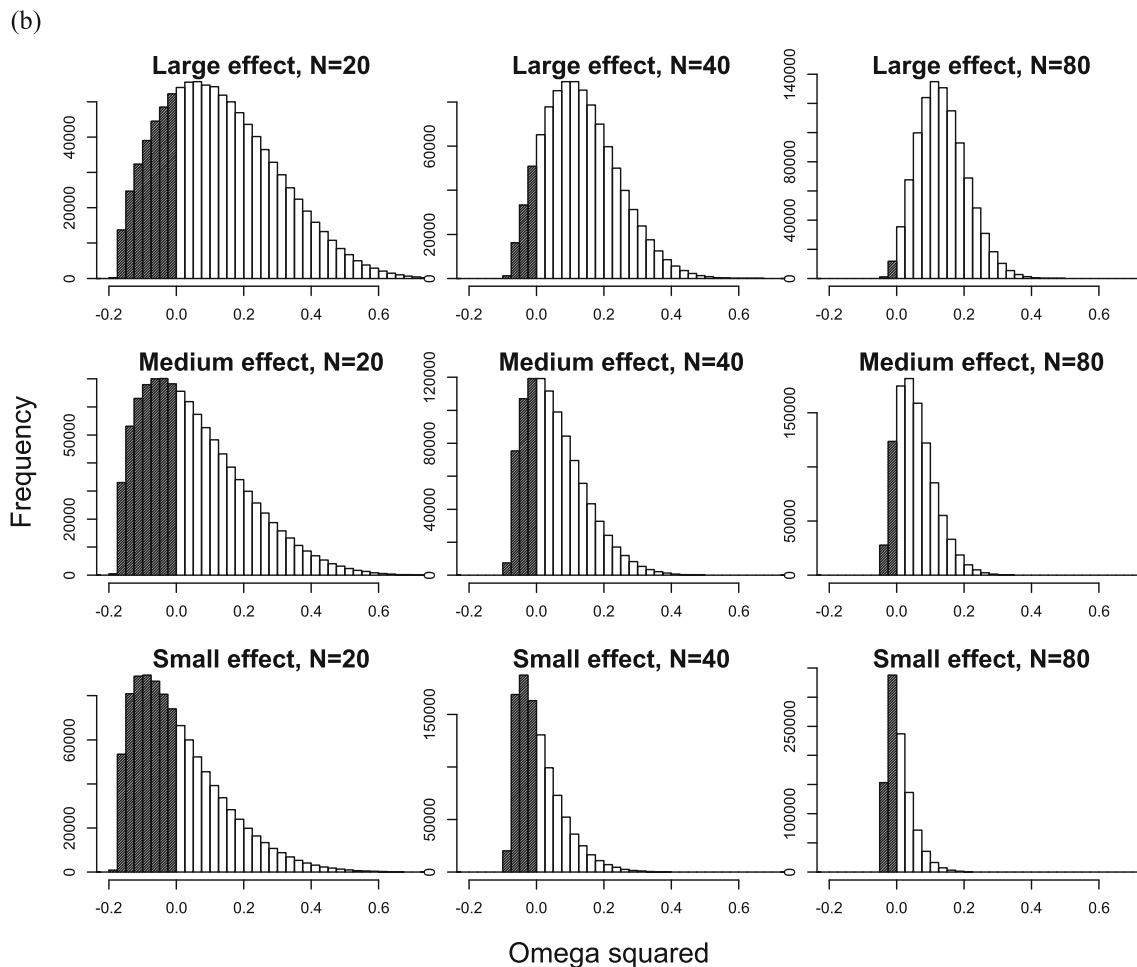


Fig. 1 (continued)

corrected estimators converge to the population quantity as the sample size increases. However, considering the empirical sample sizes used in psychology studies (Marszalek et al., 2011), the findings of the present study have some important implications for reporting practices in psychological research. Moreover, interested readers can obtain results based on other sample size conditions by simply modifying the sample size variable (nn) of the R code in the [supplementary material](#). For example, if we increase the total sample size to $N = 200$, we find that the ratios of the obtained negative estimate under intermediate mean variability are .357, .013, and .000 for small, medium, and large conditions, respectively.

Of course, no study is perfect, and this one is no exception. Although this study has focused on the point estimator of effect size, in practice, accompanying it with an interval estimate would help acknowledge the uncertainty. In particular, when a negative effect size estimate is obtained, it can mean that its sampling distribution is not narrow enough to provide a precise point estimate; thus, interval estimation may be more relevant for interpretation. For the interval estimation of effect

size, see Smithson (2001), Thompson (2007), and Finch and French (2012).

A simulation approach was employed in this study. In contrast, the probability of obtaining negative estimates in long replications can also be derived analytically. In fact, one can easily prove that this is equal to the probability that the F statistic in ANOVA will take a value less than one (Keppel & Wickens, 2004). Even so, a simulation approach is intuitive, and its results are easy to understand with graphs such as histograms and boxplots. Sample statistics such as the quantile are also easy to compute. Thanks to a million random replications per condition, the simulation-based results in this study are highly reliable.

In this study, only a simple one-factor, between-subjects design is considered. Still, the definition of population variance-accounted-for effect sizes does not depend on design, and therefore our results are expected to apply to more elaborate designs. In practice, partial effect size estimators such as partial eta squared are often reported in such designs. Partial estimates tend to take larger values (Morris & Fritz, 2013) because they partial out other factors from the

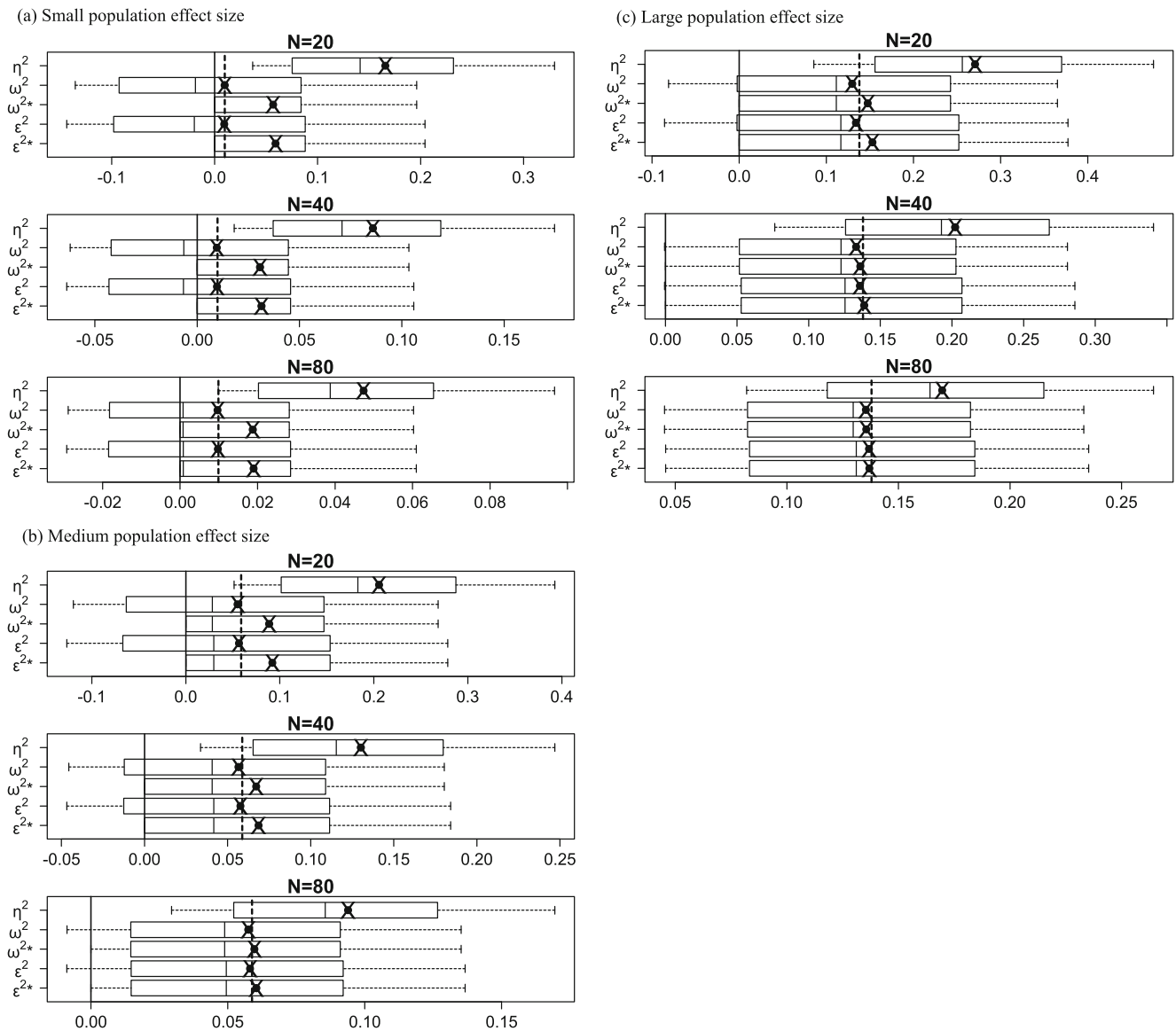


Fig. 2 Resultant boxplots of the effect size estimators for a bias comparison in the intermediate population mean variability condition. Effect sizes with an asterisk (ω^{2*} and ε^{2*}) represent cases in which negative estimates are treated as zero, whereas those without an asterisk (ω^2 and ε^2) represent cases in which negative estimates are treated as is. The η^2 statistic does not take negative values, by definition. A box indicates the interquartile range, a whisker indicates the interval

between the 10th and 90th percentiles, and the center of a cross indicates the mean of the sampling distribution. The thick vertical dashed lines indicate the actual population effect size values. The difference along the x -axis between the center of a cross and this dashed line corresponds to the bias of each estimator. Note that the range of the graphs is adjusted for each condition, to better investigate the relative performances of the estimators

denominator. Thus, the chance of obtaining negative partial effect size estimates may be smaller than in this study. Thorough investigation of partial estimates is an important direction for future study.

As the reviewers of this article pointed out, the issue of negative effect size estimates discussed here may stem from the fact that existing bias-corrected effect size statistics are, in fact, not well defined. The definitions of both ω^2 and ε^2 , as is shown in Table 1, do not include the natural constraint of positivity. Originally, ω^2 and ε^2 were designed as estimators

that correct the positive bias of η^2 . However, because the positivity constraint is not taken into account, ω^2 and ε^2 can take negative values, causing the problem discussed in this article. Therefore, exploring a variance-accounted-for effect size estimator that both is restricted to the natural range of $[0, 1]$ and corrects the bias of η^2 would be an interesting research direction for the future. A related research interest would be an exploration of the Bayesian approach. In fact, Morey et al. (2016) recently criticized the practice of arbitrarily setting the lower confidence limit of ω^2 at zero, and proposed a shift

to the Bayesian approach. Variance-accounted-for effect size is a priori known to be above zero, and this fact can be represented as a prior distribution. Nevertheless, the characteristics of Bayesian variance-accounted-for effect size estimators have not been studied well in the literature, and could be another fruitful future research direction.

Author note This work was supported by grants from the Japan Society for the Promotion of Science (24730544) and Strategic Research Foundation Grant-aided Project for Private Universities from MEXT Japan (2011–2015 S1101013).

References

- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*, 94–107. doi:10.1037/0021-9010.90.1.94
- Carroll, R. M., & Nordholm, L. A. (1975). Sampling characteristics of Kelly's ε^2 and Hays' ω^2 . *Educational and Psychological Measurement, 35*, 541–554. doi:10.1177/001316447503500304
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532–538. doi:10.1037/a0015808
- Finch, W. H., & French, B. F. (2012). A comparison of methods for estimating confidence intervals for omega-squared effect size. *Educational and Psychological Measurement, 72*, 68–77. doi:10.1177/0013164411406533
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*, 2–18. doi:10.1037/a0024338
- Glass, G. V., & Hakstian, A. R. (1969). Measures of association in comparative experiments: Their development and interpretation. *American Educational Research Journal, 6*, 403–414. doi:10.3102/00028312006003403
- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York: Routledge.
- Hays, W. L. (1963). *Statistics for psychologists* (5th ed.). New York: Holt, Rinehart & Winston.
- Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62*, 227–240. doi:10.1177/0013164402062002002
- Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences, 21*, 554–559.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River: Pearson.
- Keselman, H. J. (1975). A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Canadian Psychological Review, 16*, 44–48. doi:10.1037/h0081789
- Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington: American Psychological Association.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. doi:10.3389/fpsyg.2013.00863
- Lakens, D. (2015). Why you should use omega-squared instead of eta-squared [Blog post]. Retrieved from <http://daniellakens.blogspot.nl/2015/06/why-you-should-use-omega-squared.html>
- Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*, 612–625. doi:10.1093/hcr/28.4.612
- Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills, 112*, 331–348. doi:10.2466/03.11.PMS.112.2.331-348
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah: Erlbaum.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*, 103–123. doi:10.3758/s13423-015-0947-8
- Morris, P. E., & Fritz, C. O. (2013). Effect sizes in memory research. *Memory, 21*, 832–842. doi:10.1080/09658211.2013.763984
- Okada, K. (2016). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika, 40*, 129–147. doi:10.2333/bhmk.40.129
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286. doi:10.1006/ceps.2000.1040
- Schagen, I., & Elliot, K. (2004). *But what does it mean? The use of effect sizes in educational research*. London: National Foundation for Educational Research.
- Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior Research Methods, 45*, 536–546. doi:10.3758/s13428-012-0257-2
- Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of non-central distribution in computing intervals. *Educational and Psychological Measurement, 61*, 605–632. doi:10.1177/00131640121971392
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44*, 423–432. doi:10.1002/pits.20234
- Winkler, R. L., & Hays, W. L. (1975). *Statistics: Probability, inference, and decision* (2nd ed.). New York: Holt.