Received: 1 October 2025 | Accepted: 26 July 2026

DOI: 10.1111/bmsp.70071

the british psychological society
promoting excellence in psychology

ARTICLE

Geometric standardized mean difference and its application to meta-analysis

Jiandong Shi[1] | **Xiaochen Zhang**[2] | **Lu Lin**[3] | **Hiu Yee Kwan**[4] | **Tiejun Tong**[5]

[1]Department of Statistics and Data Science, Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, Beijing Normal-Hong Kong Baptist University, Zhuhai, China

[2]Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, China

[3]Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, China

[4]School of Chinese Medicine, Hong Kong Baptist University, Hong Kong, China

[5]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Correspondence

Xiaochen Zhang, Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, China.
Email: zhangxiaochen@bnu.edu.cn

Tiejun Tong, Department of Mathematics, Hong Kong Baptist University, Hong Kong, China.
Email: tongt@hkbu.edu.hk

Funding information

National Natural Science Foundation of China, Grant/Award Number: 12501351; the General Research Fund of Hong Kong, Grant/Award Number: HKBU12300123 and HKBU12303826; Hong Kong Baptist University, Grant/Award Number: RC-FNRA-IG/23-24/SCI/03; the Guangdong and Hong Kong Universities, Grant/Award Number: JRCS-25-26-MATH-02

Abstract

The standardized mean difference (SMD) is a widely used measure of effect size, particularly common in psychology, clinical trials and meta-analysis involving continuous outcomes. Traditionally, under the equal variance assumption, the SMD is defined as the mean difference divided by a common standard deviation. This approach is prevalent in meta-analysis but can be overly restrictive in clinical practice. To accommodate unequal variances, the conventional method averages the two variances arithmetically, which does not allow for an unbiased estimation of the SMD. Inspired by this, we propose a geometric approach to averaging the variances, resulting in a novel measure for standardizing the mean difference with unequal variances. We further propose the Cohen-type and Hedges-type estimators for the new SMD, and derive their statistical properties including the confidence intervals. Simulation results show that the Hedges-type estimator performs optimally across various scenarios, demonstrating lower bias, lower mean squared error and improved coverage probability. A real-world meta-analysis also illustrates that our new SMD and its estimators provide valuable insights to the existing literature and can be highly recommended for practical use.

KEYWORDS

Effect size, geometric averaging, meta-analysis, standardized mean difference, unbiased estimator, unequal variances

1 | INTRODUCTION

Effect size has emerged as a critical complement to statistical significance in quantitative research, reflecting its importance championed by leading scientific bodies. The American Psychological Association (Wilkinson and The Task Force on Statistical Inference, 1999) and the American Statistical Association (Wasserstein & Lazar, 2016) have advocated for effect size reporting to enrich the interpretation of research findings. Within case–control studies involving continuous outcomes, the mean difference (MD) serves as a natural and intuitive effect size metric, particularly when outcomes share identical measurement scales (Borenstein et al., 2009). This direct comparability is exemplified by universally standardized measures such as blood pressure (millimetres of mercury, mmHg), where MD yields immediately interpretable results that readily support cross-study synthesis such as the systematic review and meta-analysis.

However, the measures of an effect may not always be consistent in practice, especially in psychology. When outcome measures employ divergent scales across studies, raw MD values become fundamentally incomparable, rendering direct aggregation statistically untenable. For example, in the intelligence assessment, different studies utilize distinct psychometric instruments (e.g. Wechsler adult intelligence scale [WAIS], Stanford-Binet) that generate non-equivalent score metrics (Fancher, 1985; Kaufman & Lichtenberger, 2006), precluding meaningful MD comparison and synthesis. Another example is that the depression measurement employs disparate scales like the Beck Depression Inventory (BDI) and Hamilton Depression Rating Scale (HDRS) (Noetel et al., 2024), each with unique structural properties. The BDI's 21-item scale (0–3 per item) yields a maximum score of 63 (Beck et al., 1961), while the HDRS's variable structure (17–29 items scored 0–5) may produce maxima exceeding 80 (Hamilton, 1960). Critically, these instruments establish different clinical thresholds and numerical ranges, creating systematic bias when MDs are naively combined. In meta-analysis, studies using wider ranged scales (e.g. HDRS) would disproportionately influence pooled estimates due to their larger potential MD magnitudes, possibly yielding clinically misleading conclusions. Consequently, methodological standardization becomes essential to place all effect sizes on a common scale for the subsequent synthesis.

To deal with the problem, the standardized mean difference (SMD) is introduced in the literature as

$$\mathrm{SMD} = \frac{\text{Mean difference between two groups}}{\text{Standardizer}}, \tag{1}$$

where the standardizer is to standardize the mean difference so that the resulting SMD has a uniform scale for different studies, no matter whether the raw data are measured under different scales (Coe, 2002; Cohen, 1969). Owing to this nature, SMD perfectly matches the scope of the systematic review and meta-analysis when possibly different scales of data are collected from individual studies (Cooper et al., 2019). Naturally, SMD plays an indispensable role therein (Gallardo-Gómez et al., 2024). To further explore the standardizer in (1), let the two samples from the case and control groups, respectively, follow the normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_0, \sigma_0^2)$. When $\sigma_1^2 = \sigma_0^2 = \sigma^2$, it is natural to take σ as the standardizer, yielding the most commonly used SMD as

$$\delta = \frac{\mu_1 - \mu_0}{\sigma}. \tag{2}$$

To estimate δ which is often unknown in practice, we further let $(\overline{X}_1, S_1^2)$ and $(\overline{X}_0, S_0^2)$ be the sample means and variances of the two groups with n_1 and n_0 sample sizes, and $S_{\text{pool}} = \sqrt{\left[\left(n_1 - 1\right)S_1^2 + \left(n_0 - 1\right)S_0^2\right]/(n_1 + n_0 - 2)}$ be the pooled sample standard deviation. Cohen (1969) proposed to estimate δ by

$$d = \frac{\overline{X}_1 - \overline{X}_0}{S_{\text{pool}}}. \tag{3}$$

Cohen's d is simple and practical, yet as is well known, this estimator is often biased especially for small sample sizes. To reduce the bias, Hedges (1981) introduced a bias correction factor $J_\nu = (2/\nu)^{1/2}\Gamma(\nu/2)/\Gamma\big((\nu-1)/2\big)$, where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ is the gamma function. Under the normality assumption and equal variance, the unbiased estimator of δ, known as Hedges' g, is given as

$$g = d \cdot J_{n_1+n_0-2}. \tag{4}$$

For practical application, Hedges (1981) further applied $1 - 3/(4\nu - 1)$ to approximate J_ν with an acceptable accuracy. To conclude, with equal variance, the SMD can naturally be defined as δ in (2), and moreover, it can also be readily estimated by Cohen's d or Hedges' g. Note that Cohen's d and Hedges' g are prevalent in systematic review and meta-analysis (Higgins & Green, 2008), but the equal variance assumption can be overly restrictive in clinical practice (Hedges, 2025).

On the contrary, when the variances are not equal, the standardizer in (1) may not be immediately available for computing the SMD. To deal with this problem, a few measures have been proposed in the literature, with many of them taking a weighted arithmetic mean of the two variances in the standardizer. Consequently, the SMD with unequal variances can often be formulated as

$$\delta_w^* = \frac{\mu_1 - \mu_0}{\sqrt{w\sigma_1^2 + (1-w)\sigma_0^2}}, \tag{5}$$

where $w \in [0,1]$ is the weight assigned to the case group. Popular choices of w include, for example, 0 or 1 (Glass, 1976; Glass et al., 1981), 0.5 (Cohen, 1988) and $n_0/(n_1+n_0)$ (Aoki, 2020). The SMD proposed in Kulinskaya and Staudte (2007) and Shieh (2013) equals $\sqrt{(n_1 n_0)/(n_1+n_0)^2} \cdot \delta^*_{n_0/(n_1+n_0)}$ under our notations, rather than simply reducing to an arithmetic SMD in (5). Additional results on the estimation and inferential properties are available in the literature (Algina et al., 2006; Bird, 2023; Cumming, 2013; Delacre et al., 2021; Goulet-Pelletier & Cousineau, 2018; Huynh, 1989; Keselman et al., 2008; Peng & Chen, 2014; Suero et al., 2023). A main motivation for proposing (5) is to facilitate the estimation and inference by noting that the weighted sample variance $wS_1^2 + (1-w)S_0^2$ approximately follows a chi-squared distribution. However, this approximation leads to an obvious limitation that an exactly unbiased estimator of δ_w^* may not exist for any $w \in (0,1)$. In addition, the possibly low coverage probability of the confidence interval (CI) is also a concern. On the other hand, if we take $w = 0$ or $w = 1$, it will then suffer from another limitation that the variance from one group will be completely ignored when computing the SMD, together with possibly low coverage probability of CI again. For more details, see Section 2.

In view of the limitations on the existing SMDs, we propose a new measure for standardizing the mean difference allowing for unequal variances. Rather than to take the weighted arithmetic mean in (5), we consider a geometric approach to weight the two variances, yielding an alternative definition of SMD as

$$\delta_w = \frac{\mu_1 - \mu_0}{\sqrt{(\sigma_1^2)^w(\sigma_0^2)^{1-w}}} = \frac{\mu_1 - \mu_0}{\sigma_1^w \sigma_0^{1-w}}, \tag{6}$$

where w is the power weight assigned to the case group. It is noteworthy that the two SMDs in (5) and (6) will be the same when $w = 0$ or 1, and both of them reduce to $\delta = (\mu_1 - \mu_0)/\sigma$ as defined in (2) when the two variances are equal. With the new definition, an unbiased estimator of the new measure can be derived by a bias correction to the plug-in estimator with the mutually independent sample statistics $\overline{X}_1 - \overline{X}_0$, S_1^{-w} and $S_0^{-(1-w)}$. From this perspective, our geometric SMD in (6) may turn out to be a better choice than the arithmetic SMD in (5) for the practical implementation. We propose both Cohen-type and Hedges-type estimators for the new SMD, and derive their statistical properties including the confidence intervals. For more details, see Section 3.

The rest of the paper is organized as follows. Section 2 reviews the existing measures of SMD under unequal variances as well as presents their respective estimators. Section 3 introduces the new definition of SMD, and provides the Cohen-type and Hedges-type estimators together with their CIs. Section 4conducts simulation studies to verify that the Hedges-type estimator performs optimally across various scenarios, demonstrating lower bias, lower mean squared error and improved coverage probability. Section 5 illustrates the usefulness of our new SMD to meta-analysis. Lastly, the paper is concluded in Section 6 with discussion and the technical results are presented in the Appendices.

2 | A REVIEW OF SMDS ALLOWING FOR UNEQUAL VARIANCES

This section reviews the existing literature on standardizing the mean difference when the equal variance assumption does not hold. And as mentioned earlier, most existing SMDs can be written as a form of $\delta_w^* = (\mu_1 - \mu_0)/\sqrt{w\sigma_1^2 + (1-w)\sigma_0^2}$, with the key difference on how to assign the weight $w \in [0, 1]$.

2.1 | SMD with $w = 0$ or 1

Glass et al. (1981) recommended to take the standard deviation of the data from either group as the standardizer in (1). Without loss of generality, by taking $w = 0$, it results in the SMD as

$$\delta_0^* = \frac{\mu_1 - \mu_0}{\sigma_0}.$$

To further estimate δ_0^*, the plug-in estimator is $d_0^* = (\overline{X}_1 - \overline{X}_0)/S_0$ and the bias-corrected estimator is $g_0^* = J_{n_0-1} \cdot d_0^*$, where J_{n_0-1} is the bias correction factor as defined in Section 1. For ease of reference, we also refer to them as the Cohen-type and Hedges-type estimators, respectively. Moreover, Algina et al. (2006) showed that

$$\left(\frac{1}{n_0} + \frac{\sigma_1^2}{\sigma_0^2 n_1}\right)^{-1/2} d_0^* \sim t(n_0 - 1, \lambda),$$

where $t(n_0 - 1, \lambda)$ is the non-central t-distribution with $n_0 - 1$ degrees of freedom and non-centrality parameter $\lambda = (\mu_1 - \mu_0)/\sqrt{\sigma_1^2/n_1 + \sigma_0^2/n_0}$. Letting also $\hat{\lambda} = (\overline{X}_1 - \overline{X}_0)/\left[\sqrt{S_1^2/n_1 + S_0^2/n_0}\right]$, a $100(1-\alpha)\%$ CI of δ_0^* can be constructed as

$$\left[\sqrt{\frac{1}{n_0} + \frac{S_1^2}{S_0^2 n_1}} \cdot t_{1-\alpha/2}(n_0 - 1, \hat{\lambda}),\ \sqrt{\frac{1}{n_0} + \frac{S_1^2}{S_0^2 n_1}} \cdot t_{\alpha/2}(n_0 - 1, \hat{\lambda})\right], \tag{7}$$

where $t_{\alpha/2}(n_0 - 1, \hat{\lambda})$ represents the upper $\alpha/2$ quantile of the non-central t-distribution with $n_0 - 1$ degrees of freedom and non-centrality parameter $\hat{\lambda}$.

A main advantage of the SMD with $w = 0$ or 1 is its simplicity for practical use, together with the availability of an unbiased estimator for any fixed sample sizes. On the other hand, this simplified

SMD also has two obvious limitations. First, by taking the standard deviation of the data from only one group, the SMD cannot reflect the data variation from the other group. To be more specific, when σ_1 varies from 0.01 to 100 with μ_1, μ_0 and σ_0 fixed, δ_0 does not change anymore, which is intuitively unreasonable. Second, the $100(1-\alpha)\%$ CI in (7) may suffer from low coverage probability especially for small sample sizes. With the shorter CI, practitioners are more likely to conclude the significant result, which may be questionable.

2.2 | SMD with $w \in (0, 1)$

Other than the 0 and 1 weights, Cohen (1988) recommended to apply the equal weight $w = 0.5$, yielding $\delta^*_{0.5} = (\mu_1 - \mu_0)/\sqrt{(\sigma_1^2 + \sigma_0^2)/2}$. For more details on this SMD, one may refer to Huynh (1989) and Keselman et al. (2008). In addition, Aoki (2020) considered the weight $w = n_1^{-1}/(n_1^{-1} + n_0^{-1})$, or equivalently $w = n_0/(n_1 + n_0)$, by following a similar idea as Welch's t-test (Welch, 1938). One problem of Aoki's SMD is, however, that it will depend on the study sample sizes, which hence affects the subsequent data analysis including the meta-analytical results (Cumming, 2013).

For any SMD in this category, including $\delta^*_{0.5}$ and $\delta^*_{n_0/(n_1+n_0)}$, the Cohen-type estimator is given as

$$d^*_w = \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{wS_1^2 + (1-w)S_0^2}}. \tag{8}$$

Nevertheless, since the weighted sample variance does not follow a chi-square distribution, the Cohen-type estimator cannot yield a fully unbiased estimator even after the bias correction is implemented. As a compromise, an approximate Hedges-type estimator is known as

$$g^*_w = d^*_w \cdot J_{\hat{\nu}_w}, \tag{9}$$

where $\hat{\nu}_w = [(wS_1^2 + (1-w)S_0^2)^2]/[w^2S_1^4/(n_1-1) + (1-w)^2S_0^4/(n_0-1)]$ is the approximated degrees of freedom derived from the Welch–Satterthwaite equation (Satterthwaite, 1941; Welch, 1938). In Section 4.2, we will show that this Hedges-type estimator will always result in a non-negligible bias unless the sample sizes are sufficiently large.

Letting $\nu_w = (w\sigma_1^2 + (1-w)\sigma_0^2)^2/[w^2\sigma_1^4/(n_1-1) + (1-w)^2\sigma_0^4/(n_0-1)]$, we further show in Appendix A that, for any given $w \in (0, 1)$,

$$\left(\frac{\sigma_1^2/n_1 + \sigma_0^2/n_0}{w\sigma_1^2 + (1-w)\sigma_0^2}\right)^{-1/2} \cdot \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{wS_1^2 + (1-w)S_0^2}} \;\dot{\sim}\; t(\nu_w, \lambda), \tag{10}$$

where $\dot{\sim}$ represents 'approximately distributed as'. Moreover, by following the similar arguments as in Section 2.1, the $100(1-\alpha)\%$ CI for δ^*_w can be constructed as

$$\left[\sqrt{\frac{S_1^2/n_1 + S_0^2/n_0}{wS_1^2 + (1-w)S_0^2}} \cdot t_{1-\alpha/2}(\hat{\nu}_w, \hat{\lambda}),\ \sqrt{\frac{S_1^2/n_1 + S_0^2/n_0}{wS_1^2 + (1-w)S_0^2}} \cdot t_{\alpha/2}(\hat{\nu}_w, \hat{\lambda})\right], \tag{11}$$

which, however, also suffers from low coverage probability (Bonett, 2008). Note that for the special cases of $w = 0.5$ and $w = n_0/(n_1 + n_0)$, the corresponding results had already been derived by Aoki (2020) and Huynh (1989), and they coincide with what we derived in (10) and (11). Another way to establish the CI is by the asymptotic normality of d_w^* as well as its large sample variance calculated by the properties of the noncentral t-distribution. For the special case of $w = 0.5$, Bonett (2008) proposed an approximate normal CI with the approximated variance to improve the performance.

3 | NEW SMD AND ITS ESTIMATION

As is known, the existing SMD in (5) cannot be unbiasedly estimated for any $w \in (0, 1)$, mainly because the weighted sample variance does not follow a chi-square distribution. To overcome the problem, we now take a geometric, rather than arithmetic, approach to average the two unequal variances. More specifically, our new definition of SMD is

$$\delta_w = \frac{\mu_1 - \mu_0}{\sigma_1^w \sigma_0^{1-w}}, \tag{12}$$

where $w \in [0, 1]$ is the weight assigned to the case group, as already seen in Section 1. With the geometric weighting approach, our new SMD remains invariant for any linear transformation $f(x) = ax + b$ with $a > 0$ to both the case and control groups. Note also that our new SMD will reduce to the Glass' SMD in Section 2.1 when $w = 0$ or 1, and to the pooled SMD in (2) when $\sigma_1^2 = \sigma_0^2 = \sigma^2$. Moreover, δ_w can also be thought of as a weighted product of $\delta_1 = (\mu_1 - \mu_0)/\sigma_1$ and $\delta_0 = (\mu_1 - \mu_0)/\sigma_0$, which are two simplified SMDs that are very easy to handle. By doing so, we established a structured approach wherein the true SMD lies between δ_1 and δ_0, providing intuitive insights for interpretation. In contrast, we note that the arithmetic SMD lacks such a user-friendly interpretation since δ_w^* cannot be expressed as a form of $w\delta_1 + (1 - w)\delta_0$ accordingly.

For further comparison, the geometric SMD is always larger than or equal to the arithmetic SMD, that is, $\delta_w \geq \delta_w^*$, by noting that the weighted geometric mean is always smaller than or equal to the weighted arithmetic mean such that $(\sigma_1^2)^w(\sigma_0^2)^{1-w} \leq w\sigma_1^2 + (1 - w)\sigma_0^2$. Additionally, when $\mu_1 \neq \mu_0$, we also present their ratio $r = \delta_w/\delta_w^*$ in Figure 1 according to the degree of unbalance on the two variances, represented by $u = \log_2(\sigma_1^2/\sigma_0^2)$, together with the different choice of the weight $w \in [0, 1]$. From Figure 1, it is evident that the ratio r will never fall below 1. Also as expected, the two SMDs will provide the same measure when the two variances are equal and/or when the weight is 0 or 1.

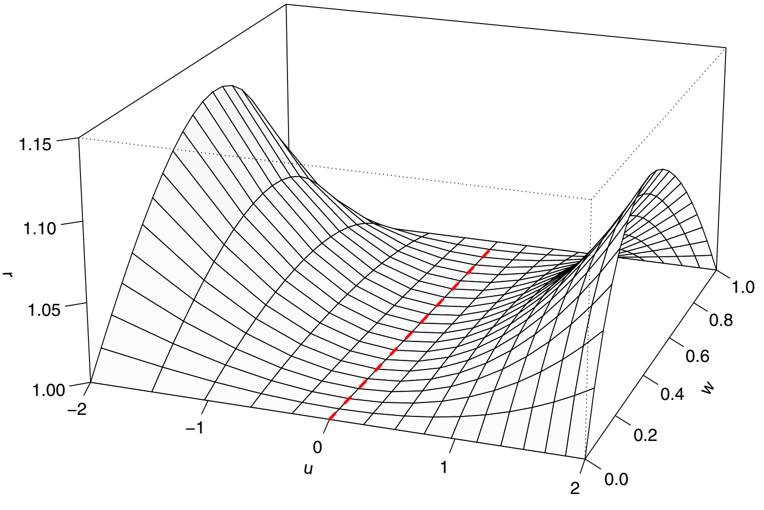

FIGURE 1 The ratio $r = \delta_w/\delta_w^*$ varies with $u = \log_2(\sigma_1^2/\sigma_0^2)$ and w. The red dashed line represents the equal variance case with $u = 0$, or equivalently, $\sigma_1^2 = \sigma_0^2$.

In contrast, the difference between the two SMDs will increase when the two variances become more unbalanced, in particular if w is also in the middle of $[0, 1]$. In such cases, the choice of standardizer should reflect the nature of the parameters being averaged. For location parameters such as the mean, the squared loss function is symmetric: deviations above and below the true value are penalized equally, making the arithmetic mean the natural choice. In contrast, for scale parameters such as the variance, the proportional (multiplicative) errors can be more meaningful, while the squared loss penalizes over-estimation much more harshly than underestimation. For example, if the true variance is 1, estimates of 0.1 and 10 are both off by a factor of 10, but the squared loss penalizes them very differently, giving values of $(0.1-1)^2 = 0.81$ and $(10-1)^2 = 81$, respectively. On the log scale, these two estimates translate to $\pm\log(10)$, which are penalized equally and treated symmetrically around $\log(1) = 0$. Therefore, using the geometric mean (i.e. averaging on the log scale) for variances or standard deviations prevents either extreme value from disproportionately influencing the standardizer, resulting in greater robustness under unequal variances. To conclude, our geometric SMD can be highly recommended as a useful alternative to the arithmetic SMD for practical use.

3.1 | Cohen-type estimation

To estimate the newly defined SMD in (12), we first apply the Cohen-type (plug-in) method owing to its simplicity. Specifically, for any given $w \in [0, 1]$, our Cohen-type estimator of δ_w is given as

$$d_w = \frac{\overline{X}_1 - \overline{X}_0}{S_1^w S_0^{1-w}}, \tag{13}$$

where $\overline{X}_1$, $\overline{X}_0$, S_1 and S_0 are the sample means and standard deviations of the two groups as in Section 1. Moreover, by Theorem 1 in Appendix B, d_w is asymptotically normally distributed with mean δ_w and variance

$$\mathrm{Var}(d_w) \approx \frac{\delta_w^2}{2}\left(\frac{w^2}{n_1} + \frac{(1-w)^2}{n_0}\right) + \frac{n_1(\sigma_0^2/\sigma_1^2)^w + n_0(\sigma_1^2/\sigma_0^2)^{1-w}}{n_1 n_0}. \tag{14}$$

In the special case when $\sigma_1^2 = \sigma_0^2$, we have $\delta_w = \delta$ so that our new d_w and Cohen's d are to estimate the same quantity. For their numerical performance, see Section 4.1. If we further take $w = n_1/(n_1 + n_0)$, then by Borenstein et al. (2009) and Hedges and Olkin (1985), the two estimators will be asymptotically the same by noting that

$$\mathrm{Var}(d) \approx \frac{\delta^2}{2(n_1 + n_0)} + \frac{n_1 + n_0}{n_1 n_0}. \tag{15}$$

More generally, when $\sigma_1^2 \neq \sigma_0^2$, the pooled SMD in (2) cannot be defined so that Cohen's d will also be meaningless. If, instead, we apply Cohen's d to estimate our newly defined SMD in (12), then as will be shown in Section 4.1, its performance will quickly get worse when the variances and/or the sample sizes become more unbalanced (Hedges, 2025).

Apart from the asymptotic variance of d_w in (14), when the sample sizes are small, Theorem 2 in Appendix C shows that the exact variance of d_w has a rather complicated form and may not be easy to implement in practice. To overcome the problem, we further show in Appendix D that the large sample formula in (14) can be, in fact, adjusted to approximate the exact variance by replacing n_1 and n_0 with $n_1 - 1$ and $n_0 - 1$, respectively. In light of this, a unified $100(1-\alpha)\%$ Cohen-type CI of δ_w for any sample sizes can then be constructed as $[d_w - z_{\alpha/2}\widehat{\mathrm{SE}}_{d_w},\ d_w + z_{\alpha/2}\widehat{\mathrm{SE}}_{d_w}]$, where

$$\widehat{\mathrm{SE}}_{d_w} = \sqrt{\frac{d_w^2}{2}\left(\frac{w^2}{n_1-1} + \frac{(1-w)^2}{n_0-1}\right) + \frac{(n_1-1)(S_0^2/S_1^2)^w + (n_0-1)(S_1^2/S_0^2)^{1-w}}{(n_1-1)(n_0-1)}} \tag{16}$$

is the estimated standard error of d_w, and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

To conclude, our Cohen-type estimator d_w provides a very simple formula for estimating the geometric SMD. Moreover, together with its well behaved $100(1-\alpha)\%$ CI, our Cohen-type estimation can be recommended for practical use in both balanced and unbalanced settings except for the extremely small sample sizes.

3.2 | Hedges-type estimation

Following the previous section, to handle the small sample sizes that may yield a non-negligible bias, we further apply the Hedges-type (bias-corrected) method to propose an unbiased estimator of the geometric SMD in (12). Specifically, following Theorem 2 that $E(d_w) = \delta_w/(B_{n_1-1,w}B_{n_0-1,1-w})$, our Hedges-type estimator of δ_w is given as

$$g_w = d_w \cdot B_{n_1-1,w}B_{n_0-1,1-w}, \tag{17}$$

where $d_w = (\overline{X}_1 - \overline{X}_0)/(S_1^w S_0^{1-w})$ is the Cohen-type estimator and $B_{\nu,w} = (2/\nu)^{w/2}\Gamma(\nu/2)/\Gamma((\nu-w)/2)$ is the bias correction factor. In addition, noting that $\mathrm{SE}(g_w) = B_{n_1-1,w}B_{n_0-1,1-w} \cdot \mathrm{SE}(d_w)$, we can further construct the $100(1-\alpha)\%$ Hedges-type CI as $[g_w - z_{\alpha/2}\widehat{\mathrm{SE}}_{g_w}, g_w + z_{\alpha/2}\widehat{\mathrm{SE}}_{g_w}]$, where

$$\widehat{\mathrm{SE}}_{g_w} = \widehat{\mathrm{SE}}_{d_w} \cdot B_{n_1-1,w}B_{n_0-1,1-w} \tag{18}$$

is the estimated standard error of g_w and $\widehat{\mathrm{SE}}_{d_w}$ is given in (16).

To promote the practical use of (17) and (18), we further provide a simple approximation for the bias correction factor as

$$B_{\nu,w} \approx 1 - \frac{(2+w)w}{4\nu-1}. \tag{19}$$

When $w=0$, we have $B_{\nu,0}=1$ so that no bias correction is needed for the related group. When $w=1$, the approximate formula reduces to $B_{\nu,1} = J_\nu \approx 1 - 3/(4\nu-1)$ as in Hedges (1981). While for other weights, we plot the true values of $B_{\nu,w}$ as well as its approximate formula on the left panels of Figure 2 for $w = 0.25$, 0.5 and 0.75 with ν up to 30, and on the right panels of Figure 2 for $\nu = 5$, 10 and 20 with $w \in [0,1]$. By the numerical results, it is evident that our formula (19) provides a very accurate approximation for the bias correction factor under all the settings, which was also supported by the error analysis as in Hedges (1981). Taking $w = 0.5$ as an example, our approximation has the maximum error within 0.002 as $\nu = 1$, within 0.0003 as $\nu \geq 10$, and within 3.3×10^{-5} as $\nu \geq 30$.

To conclude, the Hedges-type estimator further improves the Cohen-type estimator by completely eliminating the bias term. Additionally, by Gautschi's inequality (Gautschi, 1959), we can derive that $B_{\nu,w} < 1$ for all ν and $w \in (0,1]$. Thus by (18), it can be seen that the Hedges-type CI is also narrower than the Cohen-type CI.

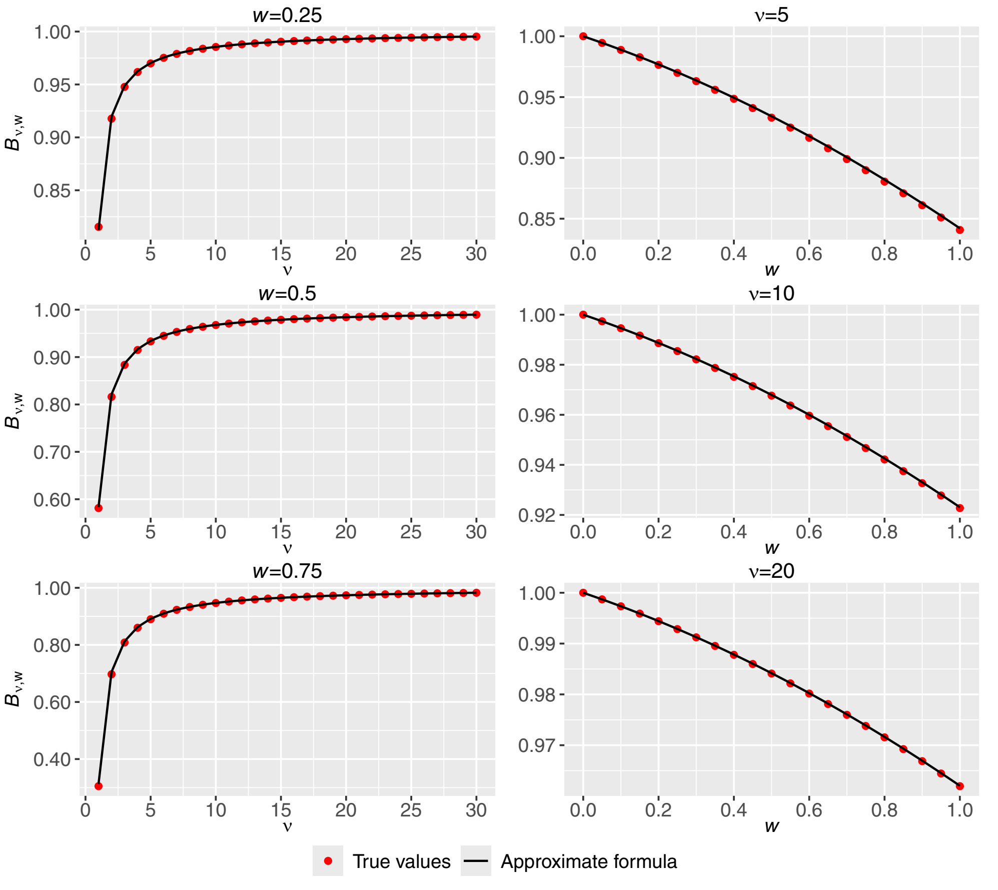

FIGURE 2 The red points represent the true values of $B_{\nu,w}$ and the black curves represent the approximate formulas in $B_{\nu,w}$ in (19). The left panels are for $w = 0.25$, 0.5 and 0.75 with ν up to 30, and the right panels are for $\nu = 5$, 10 and 20 with $w \in [0, 1]$.

4 | SIMULATION STUDIES

This section carries out simulation studies to evaluate the numerical performance of our new estimators for the geometric SMD. We also compare them with the classic estimators for the pooled SMD and the arithmetic SMD in Sections 4.1 and 4.2, respectively. To improve the traceability of our simulations, Table 1 summarizes the simulation settings for different values of w, μ_1, σ_1^2, n_1 and n_0, with the control group parameters fixed at $\mu_0 = 0$ and $\sigma_0^2 = 1$ across all scenarios. Detailed simulation settings and results are further provided in the respective subsections.

4.1 | Comparing with the pooled SMD

To compare the geometric SMD with the pooled SMD, our first simulation generates balanced data from $N(2, \sigma_1^2)$ and $N(0, \sigma_0^2)$ for the case and control groups, respectively. Without loss of generality, we set $\sigma_0^2 = 1$ and then vary σ_1^2 from $1/16$ to 16 to represent the degrees of imbalance in the variances. Or equivalently, we let $u = \log_2(\sigma_1^2/\sigma_0^2)$ range from -4 to 4, which is not an extreme setting, considering the ratios of the sample variances in our real-world meta-analysis in Section 5. For the sample sizes, we consider $n_1 = n_0 = 10$ and $n_1 = n_0 = 50$. To also explore the weight effect on the geometric SMD, we take

TABLE 1 Simulation settings for different values of w, μ_1, σ_1^2, n_1 and n_0 in this section.

Section	w	(μ_1, σ_1^2)	(n_1, n_0)
Section 4.1	$\{0.25, 0.5, 0.75\}$	$(2, \sigma_1^2), \sigma_1^2 = 1/16, 1/8, \ldots, 16$	$(10, 10), (50, 50)$
	0.5	$(2, 0.25), (2, 4)$	$(n_1, 10),\ n_1 = 5, 10, \ldots, 50$
Section 4.2	$\{0, 0.1, \ldots, 1\}$	$(2, 4)$	$(10, 10), (100, 100)$
		$(2, 1)$	$(30, 30), (150, 50)$
		$(1, 4), (0.5, 4)$	$(10, 10)$
		$(1, 1), (0.5, 1)$	$(30, 10)$

three different weights as $w = 0.25$, 0.5 and 0.75. Lastly, to compare our new estimators with Cohen's d and Hedges' g, we numerically calculate the bias, mean squared error (MSE) and coverage probability of the 95% CI for each method.

With 1,000,000 simulations, we report the bias, the MSE and the coverage probability for the four estimators in Figure 3 for $w = 0.5$, and in Appendix E for $w = 0.25$ and $w = 0.75$. From the top panels of Figure 3, when $u = 0$ so that $\sigma_1^2 = \sigma_0^2$, our Hedges-type estimator in (17) performs equally well as Hedges' g, and both of them perform better than Cohen's d and our Cohen-type estimator in (13). On the other hand, when $u \neq 0$ so that $\sigma_1^2 \neq \sigma_0^2$, Cohen's d and Hedges' g will soon get worse if we apply them to estimate The geometric SMD, since in such cases the pooled SMD cannot be defined. In contrast, our Cohen-type and Hedges-type estimators perform consistently well regardless of whether the two variances are unbalanced. Moreover, by the middle panels of Figure 3, our new estimators yield smaller MSE than Cohen's d and Hedges' g in most settings, which is more evident when u is small. In addition, as shown in the bottom panels of Figure 3, our Cohen-type and Hedges-type CIs also achieve the nominal level of the coverage probability in most settings, whereas the two classic estimators for the pooled SMD fail to do so when u is away from zero. Lastly, it is noteworthy that the above comparison results remain similar also for other weights, as displayed in Figures S1 and S2 from Appendix E for $w = 0.25$ and 0.75, respectively.

To further evaluate their numerical performance, our second simulation generates unbalanced data for the two groups with $n_0 = 10$ and n_1 varying from 5 to 50 to represent the degrees of imbalance in the sample sizes. For the variances, we consider two combinations as $(\sigma_1^2, \sigma_0^2) = (4, 1)$ and $(\sigma_1^2, \sigma_0^2) = (0.25, 1)$. All other settings are kept the same as in the first simulation including the number of simulations. We then report the bias, the MSE and the coverage probability for $w = 0.5$ in Figure 4; and to save space, the simulation results for other weights will be omitted since they are similar to those for $w = 0.5$. From the top panels of Figure 4, our Hedges-type estimator is, again, nearly unbiased in all settings. By the middle panels, it is also seen that our Hedges-type estimator provides the smallest MSE in most settings compared with the other methods. Lastly, the bottom panels show that the coverage probabilities of the 95% CIs for Cohen's d and Hedges' g are consistently lower than the nominal level. While for our new CIs, both of them yield a coverage probability very close to 95% even if the data are largely unbalanced.

4.2 | Comparing with the arithmetic SMD

As seen in Section 4.1, Cohen's d and Hedges' g only perform well when the two variances are equal. To further demonstrate the superiority of our new SMD that allows for unequal variances, we now conduct additional simulation studies to compare it with the arithmetic SMD through their respective estimators.

Specifically, our third simulation generates two groups of data with sizes n_1 and n_0 from $N(2, \sigma_1^2)$ and $N(0, \sigma_0^2)$, respectively. We also consider two distinct but complementary settings, where one is with balanced sample sizes but unequal variances ($n_1 = n_0 = 10$, $\sigma_1^2 = 4$, $\sigma_0^2 = 1$), and the other is with

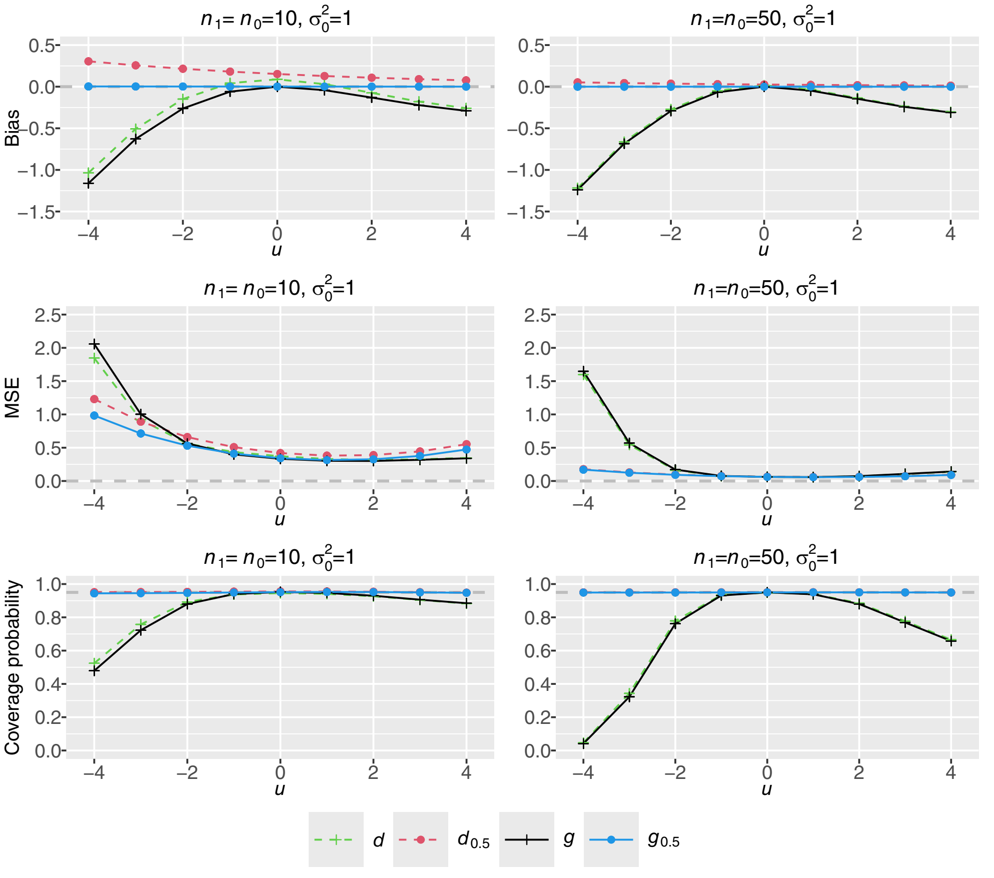

FIGURE 3 The bias, MSE and coverage probability for Cohen's d, our Cohen-type d_w with $w = 0.5$, Hedges' g, and our Hedges-type g_w with $w = 0.5$ are reported with $\sigma_0^2 = 1$ and u ranging from -4 to 4, where the left panels are for $n_1 = n_0 = 10$, and the right panels are for $n_1 = n_0 = 50$.

unbalanced sample sizes but equal variances ($n_1 = 30, n_0 = 10, \sigma_1^2 = \sigma_0^2 = 1$). Then for each $w \in [0, 1]$, we apply the estimators d_w^* in (8) and g_w^* in (9) to estimate the arithmetic SMD, and the estimators in (13) and (17) to estimate the geometric SMD, followed by their respective 95% CIs as specified. Lastly, for a fair comparison between the geometric and arithmetic SMD estimators, we numerically compute the relative bias of each method as

$$\frac{1}{T} \cdot \sum_{i=1}^{T} \frac{\text{the estimated SMD}_i - \text{the true SMD}}{\text{the true SMD}},$$

together with the relative MSE as

$$\frac{1}{T} \cdot \sum_{i=1}^{T} \left(\frac{\text{the estimated SMD}_i - \text{the true SMD}}{\text{the true SMD}} \right)^2,$$

where $T = 1,000,000$ is the number of simulations for each setting.

We report the relative bias, the relative MSE and the coverage probability for each of the four estimators in Figure 5. The top panels show that the two Cohen-type estimators both result in a significant bias. Additionally, we note that the arithmetic Hedges-type estimator also yields a non-negligible bias under

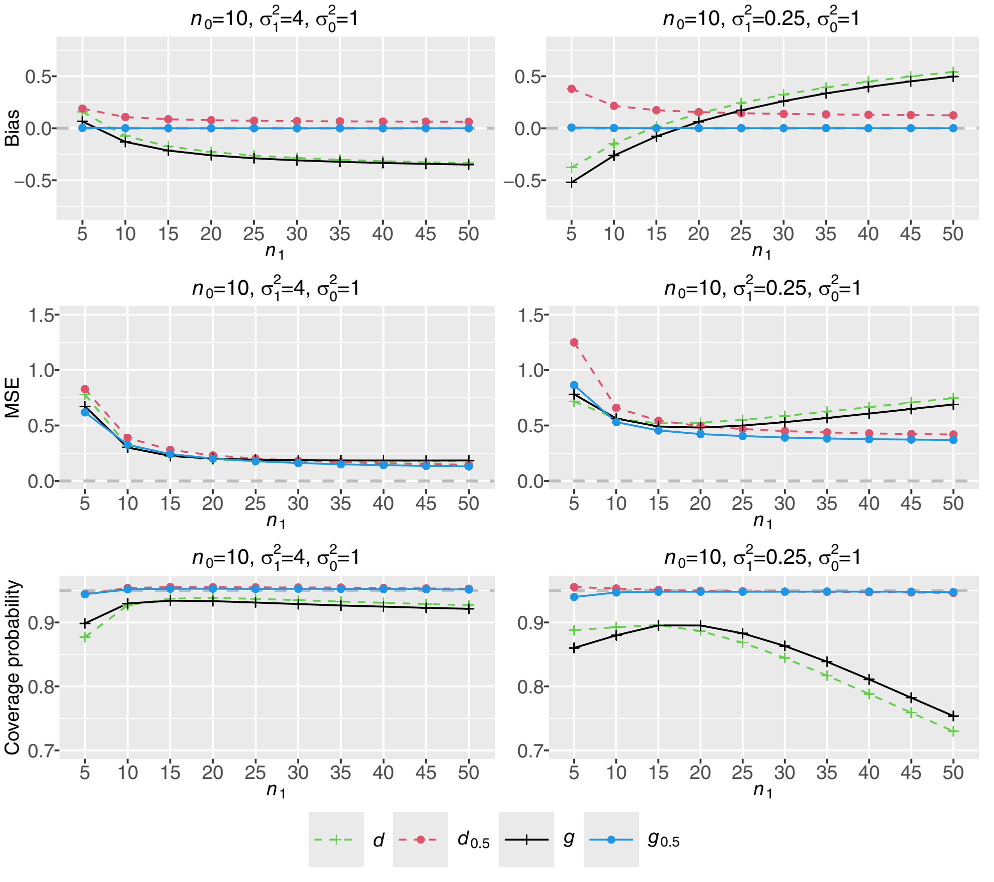

FIGURE 4 The bias, MSE and coverage probability for Cohen's d, our Cohen-type d_w with $w = 0.5$, Hedges' g, and our Hedges-type g_w with $w = 0.5$ are reported with $n_0 = 10$ and n_1 ranging from 5 to 50, where the left panels are for $\sigma_1^2 = 4$ and $\sigma_0^2 = 1$, and the right panels are for $\sigma_1^2 = 0.25$ and $\sigma_0^2 = 1$.

both scenarios, indicating that the bias correction in g_w^* is incomplete. This, in fact, coincides with our derivation in (10) that d_w^* does not exactly follow the non-central t-distribution with the degrees of freedom as $\nu_w = [(w\sigma_1^2 + (1-w)\sigma_0^2)^2]/[w^2\sigma_1^4/(n_1-1) + (1-w)^2\sigma_0^4/(n_0-1)]$, and subsequently the bias correction term J_{ν_w} cannot fully eliminate the bias. Moreover, with ν_w unknown in practice, one needs to further apply the plug-in estimate $\hat{\nu}_w = [(wS_1^2 + (1-w)S_0^2)^2]/[w^2S_1^4/(n_1-1) + (1-w)^2S_0^4/(n_0-1)]$, which also introduces new errors. On the other hand, with the geometric weighting approach, our new estimator g_w is able to completely eliminate the bias associated with the Cohen-type estimator under any combination of the weight, the variances and the sample sizes. Note that the superiority of our geometric Hedges-type estimator is also evidenced by the middle panels of Figure 5 from the perspective of the relative MSE. From the bottom panels of Figure 5, it is clear that the two CIs for the arithmetic SMD are not able to give 95% coverage probability. In contrast, both our Cohen-type and Hedges-type CIs can provide a coverage probability very close to the nominal level, especially when the weight is in the middle of [0, 1]. Following the discussion at the end of Section 3.2, our Hedges-type CI is narrower than our Cohen-type CI, which explains why its coverage probability is also slightly lower than that of its counterpart.

More simulation settings for larger sample sizes ($n_1 = n_0 = 100$ and $n_1 = 150$, $n_0 = 50$) and for different true effects by varying μ_1 as 1 and 0.5 are considered with the results shown in Figures S3–S5 of Appendix E. It is as expected that the Cohen-type estimation performs more similarly to the

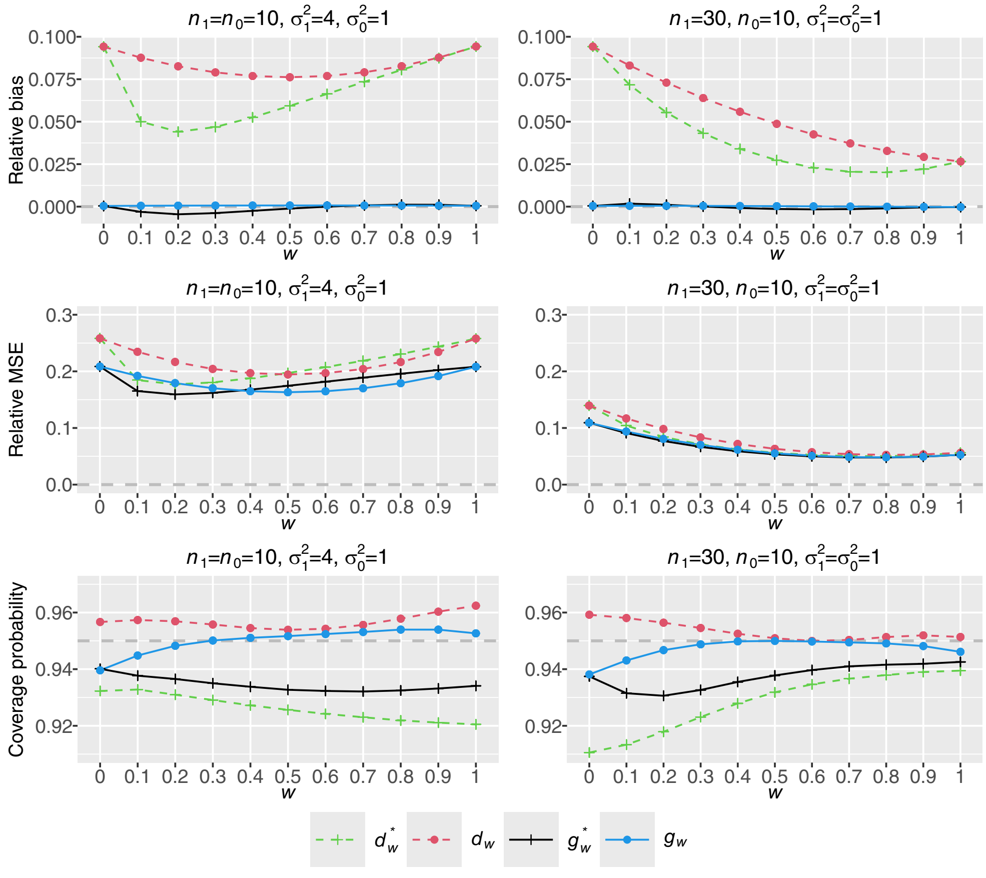

FIGURE 5 The relative bias, relative MSE, coverage probability for arithmetic Cohen-type d_w^*, our Cohen-type d_w, arithmetic Hedges-type g_w^* and our Hedges-type g_w with $w \in [0, 1]$ are reported, where the left panels are for balanced sample sizes but unequal variances, and the right panels are for unbalanced sample sizes but equal variances.

Hedges-type estimation as the sample sizes increase, while the results remain similar with the change of the true effect. When $w = 0$, both the arithmetic and geometric SMDs reduce to Glass' SMD, that is, $\delta_0^* = \delta_0$; hence, their corresponding Hedges-type estimators g_0^* and g_0 are indeed estimating Glass' SMD. Following the above simulation results, g_0^* and g_0 perform similarly well in terms of the bias and MSE, although g_0^* exhibits slightly reduced coverage probability, especially when the mean difference is small. It is also worth noting in most settings that both our Cohen-type and Hedges-type estimators with $w = 0.5$ perform better than those with other weights in terms of MSE and coverage probability.

To sum up, when the assumption of equal variance is violated, the pooled SMD in (2) becomes invalid so that the well-known Cohen's d and Hedges' g can also be misleading. Furthermore, the arithmetic SMD has its limitations in the sense that an unbiased estimate is hard to obtain, in addition to its complex form which may also hinder its real application. As an alternative, our newly introduced geometric SMD, along with the Cohen-type and Hedges-type estimators, brings new insights to the existing literature and can thus be highly recommended for practical applications.

5 | APPLICATION TO META-ANALYSIS

When embedded in meta-analysis, the geometric SMD, like the classic estimators such as Cohen's d and Hedges' g, fundamentally assumes the associated estimators (i.e. the estimated effect sizes) to

Study	Treatment Mean	Treatment SD	Treatment Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1. Spector et al. 2003 (dementia, QoL–AD)	1.30	5.10	115	−0.80	5.60	86	2.9%	0.39 [0.11; 0.68]
2. Coen et al. 2011 (dementia, QoL–AD)	3.60	3.70	14	0.50	4.40	13	2.7%	0.74 [−0.04; 1.53]
3. Alves et al. 2014 (MCI and dementia, QoL–AD)	−2.30	5.21	10	0.57	4.35	7	2.5%	−0.56 [−1.55; 0.43]
4. Cove et al. 2014 (dementia, QoL–AD, CS)	−0.77	0.39	21	0.54	0.08	21	2.4%	−4.57 [−5.75; −3.38]
5. Orrell et al. 2014 (dementia, QoL–AD, 3 months)	−1.81	6.20	114	−2.53	4.91	104	2.9%	0.13 [−0.14; 0.39]
6. Orrell et al. 2014 (dementia, QoL–AD, 6 months)	−0.48	9.92	106	1.66	9.05	93	2.9%	−0.22 [−0.50; 0.06]
7. Orrell et al. 2014 (dementia, DEMQOL, 3 months)	−4.95	14.08	114	−4.39	12.67	104	2.9%	−0.04 [−0.31; 0.22]
8. Orrell et al. 2014 (dementia, DEMQOL, 6 months)	−5.67	25.65	106	6.27	22.63	93	2.9%	−0.49 [−0.77; −0.21]
9. Orrell et al. 2014 (dementia, QoL–AD proxy, 3 months)	0.23	5.31	114	−0.90	6.01	104	2.9%	0.20 [−0.07; 0.47]
10. Orrell et al. 2014 (dementia, QoL–AD proxy, 6 months)	0.42	8.62	106	0.75	8.70	93	2.9%	−0.04 [−0.32; 0.24]
11. Orrell et al. 2014 (dementia, DEMQOL proxy, 3 months)	−0.84	28.51	114	−4.08	27.23	104	2.9%	0.12 [−0.15; 0.38]
12. Orrell et al. 2014 (dementia, DEMQOL proxy, 6 months)	−4.45	15.01	106	−5.59	16.03	93	2.9%	0.07 [−0.21; 0.35]
13. Orgeta et al. 2015 (dementia, QoL–AD, 13 weeks)	−0.11	0.60	142	0.13	0.41	146	2.9%	−0.47 [−0.70; −0.23]
14. Orgeta et al. 2015 (dementia, QoL–AD, 26 weeks)	0.42	0.19	134	−0.25	0.13	139	2.8%	4.12 [3.70; 4.54]
15. Orgeta et al. 2015 (dementia, QoL–AD proxy, 13 weeks)	0.24	0.58	142	−1.16	0.38	146	2.8%	2.86 [2.53; 3.19]
16. Orgeta et al. 2015 (dementia, QoL–AD proxy, 26 weeks)	0.42	0.63	134	−1.10	0.08	139	2.8%	3.41 [3.03; 3.78]
17. Orgeta et al. 2015 (dementia, DEMQOL, 13 weeks)	0.23	0.84	142	1.87	1.75	146	2.9%	−1.19 [−1.44; −0.94]
18. Ortega et al. 2015 (dementia, DEMQOL, 26 weeks)	1.61	0.59	134	2.94	2.44	139	2.9%	−0.74 [−0.99; −0.50]
19. Ortega et al. 2015 (dementia, DEMQOL proxy, 13 weeks)	1.27	0.79	142	0.16	0.80	146	2.9%	1.39 [1.13; 1.65]
20. Orgeta et al. 2015 (dementia, DEMQOL proxy, 26 weeks)	1.43	0.76	134	−0.41	0.04	139	2.8%	3.44 [3.07; 3.82]
21. Capotosto et al. 2017 (dementia, QoL–AD)	1.25	0.07	20	0.05	0.45	19	2.5%	3.70 [2.63; 4.77]
22. Djabelkhir et al. 2017 (MCI, EQVPA)	0.10	2.30	9	0.30	1.60	10	2.6%	−0.10 [−1.00; 0.80]
23. Piras et al. 2017 (dementia, QoL–AD)	2.30	0.37	21	−0.43	0.95	14	2.4%	4.03 [2.82; 5.23]
24. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–VAS)	63.17	29.02	12	43.52	61.19	6	2.5%	0.45 [−0.55; 1.44]
25. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–Index Value)	0.59	0.75	12	0.61	0.73	6	2.5%	−0.03 [−1.01; 0.95]
26. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–Index Value)	0.59	0.66	12	0.76	0.40	6	2.5%	−0.28 [−1.27; 0.70]
27. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–VAS)	51.83	45.61	12	42.94	52.45	6	2.5%	0.18 [−0.81; 1.16]
28. Folkerts et al. 2018 (dementia, QUALIDEM)	18.60	9.51	12	15.78	3.78	6	2.5%	0.33 [−0.66; 1.32]
29. Leroi et al. 2019 (PD–MCI, PDD, DLB, PDQ–39)	−1.72	3.66	17	−0.31	0.93	23	2.7%	−0.56 [−1.20; 0.08]
30. Leroi et al. 2019 (PD–MCI, PDD, DLB, EuroQoL–5D–Index)	0.06	0.01	24	0.02	0.01	32	2.6%	3.94 [3.02; 4.87]
31. Leroi et al. 2019 (PD–MCI, PDD, DLB, EuroQoL–5D–VAS)	−0.66	4.23	24	−6.75	4.82	31	2.7%	1.31 [0.72; 1.90]
32. Lok et al. 2019 (dementia, QoL–AD)	12.30	3.11	30	−1.47	0.98	30	2.4%	5.89 [4.69; 7.10]
33. Tsai et al 2019 (MCI and dementia, QoL–AD)	1.42	0.41	12	0.15	0.24	13	2.3%	3.70 [2.34; 5.06]
34. Alvares–Pereira et al. 2020 (dementia, QoL–AD)	−0.04	3.72	55	−0.51	4.48	50	2.8%	0.11 [−0.27; 0.50]
35. Gibbor et al. 2020 (dementia, QoL–AD)	2.09	0.59	17	0.93	0.60	16	2.6%	1.90 [1.06; 2.74]
36. Carbone et al. 2021 (dementia, QoL–AD, post–intervention)	1.52	2.08	123	−0.31	0.10	101	2.8%	1.18 [0.90; 1.47]
37. Carbone et al. 2021 (dementia, QoL–AD, 23 weeks)	1.09	2.03	106	−0.49	0.17	80	2.8%	1.02 [0.71; 1.33]
Total (95% CI)			**2662**			**2504**	**100.0%**	**0.93 [−0.03; 1.89]**

Heterogeneity: $Tau^2 = 1.8708$; $Chi^2 = 1859.43$, df = 36 (P = 0); $I^2 = 98\%$

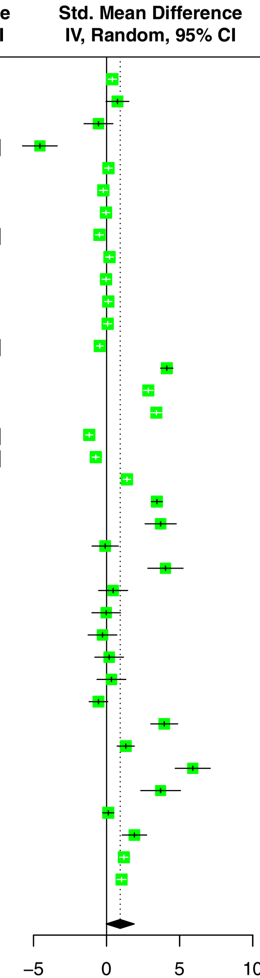

FIGURE 6 The original meta-analysis results in Gómez-Soria et al. (2023).

Study	Treatment Mean	Treatment SD	Treatment Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1. Spector et al. 2003 (dementia, QoL–AD)	1.30	5.10	115	−0.80	5.60	86	2.8%	0.39 [0.11; 0.67]
2. Coen et al. 2011 (dementia, QoL–AD)	3.60	3.70	14	0.50	4.40	13	2.7%	0.73 [−0.05; 1.51]
3. Alves et al. 2014 (MCI and dementia, QoL–AD)	−2.30	5.21	10	0.57	4.35	7	2.6%	−0.55 [−1.50; 0.40]
4. Cove et al. 2014 (dementia, QoL–AD, CS)	−0.77	0.39	21	0.54	0.08	21	2.5%	−7.18 [−9.03; −5.34]
5. Orrell et al. 2014 (dementia, QoL–AD, 3 months)	−1.81	6.20	114	−2.53	4.91	104	2.8%	0.13 [−0.14; 0.40]
6. Orrell et al. 2014 (dementia, QoL–AD, 6 months)	−0.48	9.92	106	1.66	9.05	93	2.8%	−0.22 [−0.50; 0.05]
7. Orrell et al. 2014 (dementia, DEMQOL, 3 months)	−4.95	14.08	114	−4.39	12.67	104	2.8%	−0.04 [−0.31; 0.22]
8. Orrell et al. 2014 (dementia, DEMQOL, 6 months)	−5.67	25.65	106	6.27	22.63	93	2.8%	−0.49 [−0.77; −0.21]
9. Orrell et al. 2014 (dementia, QoL–AD proxy, 3 months)	0.23	5.31	114	−0.90	6.01	104	2.8%	0.20 [−0.07; 0.47]
10. Orrell et al. 2014 (dementia, QoL–AD proxy, 6 months)	0.42	8.62	106	0.75	8.70	93	2.8%	−0.04 [−0.32; 0.24]
11. Orrell et al. 2014 (dementia, DEMQOL proxy, 3 months)	−0.84	28.51	114	−4.08	27.23	104	2.8%	0.12 [−0.15; 0.38]
12. Orrell et al. 2014 (dementia, DEMQOL proxy, 6 months)	−4.45	15.01	106	−5.59	16.03	93	2.8%	0.07 [−0.21; 0.35]
13. Orgeta et al. 2015 (dementia, QoL–AD, 13 weeks)	−0.11	0.60	142	0.13	0.41	146	2.8%	−0.48 [−0.72; −0.24]
14. Orgeta et al. 2015 (dementia, QoL–AD, 26 weeks)	0.42	0.19	134	−0.25	0.13	139	2.8%	4.24 [3.81; 4.68]
15. Orgeta et al. 2015 (dementia, QoL–AD proxy, 13 weeks)	0.24	0.58	142	−1.16	0.38	146	2.8%	2.97 [2.63; 3.31]
16. Orgeta et al. 2015 (dementia, QoL–AD proxy, 26 weeks)	0.42	0.63	134	−1.10	0.08	139	2.8%	6.74 [6.00; 7.48]
17. Orgeta et al. 2015 (dementia, DEMQOL, 13 weeks)	0.23	0.84	142	1.87	1.75	146	2.8%	−1.35 [−1.63; −1.06]
18. Ortega et al. 2015 (dementia, DEMQOL, 26 weeks)	1.61	0.59	134	2.94	2.44	139	2.8%	−1.10 [−1.46; −0.74]
19. Ortega et al. 2015 (dementia, DEMQOL proxy, 13 weeks)	1.27	0.79	142	0.16	0.80	146	2.8%	1.39 [1.13; 1.65]
20. Orgeta et al. 2015 (dementia, DEMQOL proxy, 26 weeks)	1.43	0.76	134	−0.41	0.04	139	2.8%	10.50 [9.35; 11.66]
21. Capotosto et al. 2017 (dementia, QoL–AD)	1.25	0.07	20	0.05	0.45	19	2.6%	6.53 [4.65; 8.41]
22. Djabelkhir et al. 2017 (MCI, EQVPA)	0.10	2.30	9	0.30	1.60	10	2.6%	−0.10 [−1.02; 0.82]
23. Piras et al. 2017 (dementia, QoL–AD)	2.30	0.37	21	−0.43	0.95	14	2.5%	4.42 [3.02; 5.82]
24. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–VAS)	63.17	29.02	12	43.52	61.19	6	2.6%	0.42 [−0.80; 1.65]
25. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–Index Value)	0.59	0.75	12	0.61	0.73	6	2.6%	−0.02 [−0.98; 0.93]
26. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–Index Value)	0.59	0.66	12	0.76	0.40	6	2.6%	−0.30 [−1.23; 0.63]
27. Folkerts et al. 2018 (dementia, EuroQoL–5D–5L–VAS)	51.83	45.61	12	42.94	52.45	6	2.6%	0.16 [−0.83; 1.16]
28. Folkerts et al. 2018 (dementia, QUALIDEM)	18.60	9.51	12	15.78	3.78	6	2.6%	0.43 [−0.57; 1.43]
29. Leroi et al. 2019 (PD–MCI, PDD, DLB, PDQ–39)	−1.72	3.66	17	−0.31	0.93	23	2.7%	−0.74 [−1.71; 0.24]
30. Leroi et al. 2019 (PD–MCI, PDD, DLB, EuroQoL–5D–Index)	0.06	0.01	24	0.02	0.01	32	2.6%	3.90 [2.99; 4.82]
31. Leroi et al. 2019 (PD–MCI, PDD, DLB, EuroQoL–5D–VAS)	−0.66	4.23	24	−6.75	4.82	31	2.7%	1.32 [0.73; 1.90]
32. Lok et al. 2019 (dementia, QoL–AD)	12.30	3.11	30	−1.47	0.98	30	2.5%	7.72 [6.16; 9.27]
33. Tsai et al 2019 (MCI and dementia, QoL–AD)	1.42	0.41	12	0.15	0.24	13	2.5%	3.83 [2.44; 5.21]
34. Alvares–Pereira et al. 2020 (dementia, QoL–AD)	−0.04	3.72	55	−0.51	4.48	50	2.8%	0.11 [−0.27; 0.50]
35. Gibbor et al. 2020 (dementia, QoL–AD)	2.09	0.59	17	0.93	0.60	16	2.7%	1.87 [1.05; 2.69]
36. Carbone et al. 2021 (dementia, QoL–AD, post–intervention)	1.52	2.08	123	−0.31	0.10	101	2.8%	3.99 [3.10; 4.88]
37. Carbone et al. 2021 (dementia, QoL–AD, 23 weeks)	1.09	2.03	106	−0.49	0.17	80	2.8%	2.67 [1.96; 3.39]
Total (95% CI)			**2662**			**2504**	**100.0%**	**1.35 [−0.19; 2.89]**

Heterogeneity: $Tau^2 = 2.4391$; $Chi^2 = 1949.73$, df = 36 (P = 0); $I^2 = 98\%$

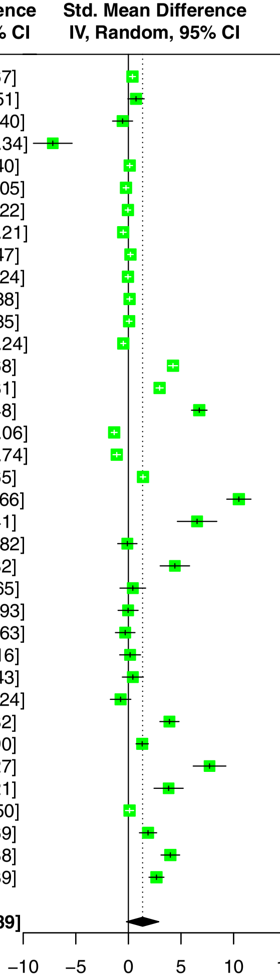

FIGURE 7 The new results by our geometric SMD and its Hedges-type estimation.

be normally distributed. Theorem 1 of Appendix B has established the asymptotic normality for our Cohen-type estimator, which can be extended to the Hedges-type estimator. It has provided the theoretical basis for the normality assumption of our new estimators for its use within the meta-analytic models.

To further illustrate the usefulness of the new SMD and its estimation, we present a real data example in meta-analysis. Gómez-Soria et al. (2023) carried out a systematic review and several meta-analyses to evaluate the effect of the cognitive stimulation (CS) on psychosocial results in older adults. They proposed to demonstrate that the quality of life (QoL), depression, anxiety and activities of daily living can be improved independent of the pharmacological treatment such as acetylcholinesterase inhibitors by CS.

In this section, we focus on the meta-analysis for QoL and report its original results in Figure 6. Gómez-Soria et al. (2023) used the SMD to measure the effect in each study and applied Hedges' g to estimate the SMD. To accommodate studies with data collected at different follow-up times, the meta-analysis employed the robust variance estimator by Tipton and Pustejovsky (2015) and applied the Satterthwaite adjustment to the degrees of freedom in the random-effects model (Borenstein et al., 2009). It is noted from the forest plot in Figure 6 that the overall SMD estimate is 0.93 with the 95% CI being $[-0.03, 1.89]$, which shows an insignificant result in favor of the case group. The lower bound of the 95% CI is very close to 0 and the p-value is 0.056, which causes difficulty in making a convincing conclusion. We note that the unequal variances appear in many studies included in the meta-analysis. Taking the 20th study as an example, the sample standard deviation from the case group is 0.76, which is 19 times of the sample standard deviation 0.04 from the control group. Moreover, there are 13 (out of a total of 37) studies with the ratios of sample standard deviations between the case and control groups outside the interval $[0.5, 2]$. This indicates that the pooled SMD in (2) may no longer be appropriate as the true measure, and so is Hedges' g to estimate the SMD.

To handle the unequal variances, we apply both the arithmetic and geometric SMDs with $w = 0.5$, as well as their Hedges-type estimates, for all 37 studies included in the meta-analysis. For the arithmetic SMD, we note that the SMD estimates for the individual studies do not differ much from those from Hedges' g. This is mainly because of the approximate equivalence between g and $g^*_{0.5}$ for the studies with nearly balanced sample sizes, regardless of whether or not the sample standard deviations largely differ. As a consequence, the meta-analytical results remain similar with the overall SMD estimate being 0.9 and the 95% CI being $[-0.04, 1.83]$. Moreover, the p-value for the overall SMD estimate is 0.057, which also suffers from the difficulty in making a convincing conclusion. For more details, please refer to the forest plot in Figure S6 of Appendix F.

In contrast, our geometric SMD significantly alters the individual SMD estimates, which, in turn, has a notable impact on the overall meta-analytical results. From Figure 7 for our new results, we observe that the SMD estimate in the 20th study is 10.5 with the 95% CI being $[9.35, 11.66]$ while the original results in Figure 6 show the SMD estimate as 3.44 with 95% CI being $[3.07, 3.82]$. This also occurs similarly for other studies with unequal sample standard deviations, for example, studies 4, 16, 21 and so on. Consequently, the meta-analysis leads to the overall SMD estimate as 1.35 with the 95% CI being $[-0.19, 2.89]$, and the p-value is 0.076. Based on the new results, it is more convincing to draw the conclusion of an insignificant result in favor of the case group. Taken together, it thus offers fresh perspectives for our new method to serve as a valuable alternative to Hedges' g for practical use.

6 | CONCLUSIONS AND DISCUSSION

For clinical studies with continuous outcomes, the standardized mean difference (SMD) is the most widely used effect size that can measure the raw data of different studies onto the same scale. Under the equal variance assumption, the pooled SMD is clearly defined as the mean difference divided by the common standard deviation, together with Cohen's d and Hedges' g as the popular estimates. In many clinical studies, however, the equal variance assumption may not be realistic, presenting a significant limitation for the practical use of Cohen's d and Hedges' g (Hedges, 2025). To deal with the unequal variances, several definitions for standardizing the mean difference had been introduced in the previous literature, with the essential idea being to take the arithmetic mean of the two variances for standardization, referred to as the arithmetic SMD. Nevertheless, as has been extensively discussed, this measure has serious limitations including, but not limited to, the complex estimation form, unavoidable bias and low coverage probability. These limitations restrict its practical utility, leaving Cohen's d and Hedges' g as the default choices even with largely unbalanced variances.

One main contribution of this paper is to introduce a new definition of SMD as $\delta_w = (\mu_1 - \mu_0)/(\sigma_1^w \sigma_0^{1-w})$ by taking the geometric mean of the two variances for standardization. To the best of our limited knowledge, this is the first work to consider the geometric weighting for the standardizer in (1), which makes our new SMD very unique compared to the conventional arithmetic SMD. We further develop the easy-to-implement estimators of the geometric SMD using the Cohen-type and Hedges-type methods. Notably, the geometric weighting approach enables an unbiased Hedges-type estimator, which is a critical advantage unattainable with the arithmetic SMD. Simulation results show that our new estimators are superior to the existing ones under a wide range of settings in terms of the relative bias, relative MSE and coverage probability. We also apply our new SMD to a real-world meta-analysis involving many studies with unequal variances, which results in more convincing results compared to the original results with Hedges' g as the effect size. To conclude, our geometric SMD does not impose the equal variance assumption as in the pooled SMD, and moreover, it has some nice properties and advantages that the arithmetic SMD does not have. And accordingly, as the effect size for meta-analysis, our Cohen-type and Hedges-type estimators for the new SMD can also be highly recommended to supersede the original Cohen's d and Hedges' g.

From a practical perspective, the adoption of our geometric SMD offers two key advantages for clinicians and researchers. First, it provides a more reliable measure of effect size when comparing treatments across studies with unequal variances, thereby increasing confidence in the conclusions drawn from systematic reviews and meta-analysis. Second, the superior statistical properties of our Hedges-type estimators, that is, the unbiasedness and more accurate CI, enable a more precise interpretation of treatment effects. This is particularly valuable in fields like psychological research, where different measurement scales (e.g. BDI vs HDRS for depression) are common and the unequal variances are frequently observed.

At the same time, several interesting directions remain open and warrant further work. First, although our findings indicate that the geometric SMD outperforms the arithmetic SMD in terms of estimation, the choice between geometric and arithmetic standardization may in practice be guided by the data characteristics and the specific context of application. Both metrics represent alternative approaches to standardizing the mean difference, yet without the equal variance assumption, neither is as immediately intuitive or interpretable as the conventional SMD in (2). Comprehensive exploration and systematic comparisons across different SMDs may yield more suitable effect size measures for various practical scenarios.

Second, a key methodological consideration for both geometric and arithmetic SMDs is the selection of the weight w. For practical use, choosing an equal weight ($w = 0.5$) serves as a reasonable initial choice for our geometric SMD. This is not only due to its simple and interpretable form as $\delta_{0.5} = (\mu_1 - \mu_0)/\sqrt{\sigma_1 \sigma_0}$, but also considering its satisfactory performance in estimation across various settings as demonstrated in Section 4. However, determining the optimal weight is a complex issue that we cannot fully resolve within the current work, and it remains a substantive topic for future research.

Third, it is also of particular importance to consider the choice between the Cohen-type estimators (including Cohen's d) and the Hedges-type estimators (including Hedges' g). Although Cohen's d is widely recognized and easy to interpret, Hedges' g is generally preferred for point estimation of SMD in individual studies due to its elegant bias correction, especially for small sample sizes. This rationale also applies to our newly proposed methodology; that is, our Hedges-type estimator in (17) can be highly recommended for practical use in estimating the geometric SMD in individual studies. Nevertheless, in the context of meta-analysis, a recent work by Lin and Aloe (2021) observed that using Hedges' g may result in a more negatively biased overall effect estimate than using Cohen's *d*. This finding highlights the need for future research to comprehensively compare the Cohen-type and Hedges-type estimators across different settings and criteria.

Lastly, we highlight the importance to distinguish between the true unequal variance case and the sample variance difference due to outliers. In practice, when raw data are available, researchers are suggested to carefully screen the data and investigate whether extreme values are attributable to measurement errors, recording errors or other data-quality problems. Such errors should be corrected whenever possible or excluded based on transparent and well-justified criteria. Sensitivity analyses including and excluding the affected data may also be conducted to evaluate their impact on the estimated SMD. When only summary statistics are available, individual-level outliers or measurement errors typically cannot be directly identified. In such cases, researchers may conduct sensitivity analyses to assess the influence of studies that potentially contain outliers. Further investigation in this area is warranted.

In conclusion, while the geometric SMD provides a valuable addition to the toolbox of effect size metrics, its implementation necessitates careful consideration of the context and characteristics of the data. Further exploration into refining the interpretation of geometric SMD, alongside understanding its assumptions, robustness and behaviour under different situations by theoretical and/or numerical evaluations, will significantly contribute to the discipline and support practitioners in making informed decisions.

AUTHOR CONTRIBUTIONS

Jiandong Shi: conceptualization; methodology; investigation; validation; formal analysis; writing – original draft; writing – review and editing. **Xiaochen Zhang:** conceptualization; methodology; investigation; validation; formal analysis; funding acquisition; writing – original draft; writing – review and editing. **Lu Lin:** methodology; supervision; writing – review and editing. **Hiu Yee Kwan:** writing – review and editing; supervision. **Tiejun Tong:** conceptualization; methodology; validation; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

The authors thank the editor, associate editor and two reviewers for their constructive comments and suggestions to greatly improve the quality of the manuscript.

FUNDING INFORMATION

Xiaochen Zhang's research was supported by National Natural Science Foundation of China (12501351). Tiejun Tong's research was supported by the General Research Fund of Hong Kong (HKBU12300123 and HKBU12303826), the Initiation Grant for Faculty Niche Research Areas of Hong Kong Baptist University (RC-FNRA-IG/23-24/SCI/03), and the Guangdong and Hong Kong Universities '1+1+1' Joint Research Collaboration Scheme (JRCS-25-26-MATH-02).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

DISCLOSURE OF ARTIFICIAL INTELLIGENCE-GENERATED CONTENT (AIGC) TOOLS

The artificial intelligence tools have only been used to review and check the writing.

ORCID

Xiaochen Zhang https://orcid.org/0000-0002-0300-1709

REFERENCES

Algina, J., Keselman, H. J., & Penfield, R. D. (2006). Confidence intervals for an effect size when variances are not equal. *Journal of Modern Applied Statistical Methods*, *5*, 2–13.

Aoki, S. (2020). Effect sizes of the differences between means without assuming variance equality and between a mean and a constant. *Heliyon*, *6*, e03306.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Bird, K. D. (2023). Improved confidence intervals for differences between standardized effect sizes. *Psychological Methods*, *28*, 1142–1153.

Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, *13*, 99–109.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley and Sons.

Coe, R. (2002). It's the effect size, stupid. *Paper presented at the British Educational Research Association Annual Conference*, Exeter, September 12–14, 2002.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

Cumming, G. (2013). Cohen's d needs to be readily interpretable: comment on Shieh (2013). *Behavior Research Methods*, *45*, 968–971.

Delacre, M., Lakens, D., Ley, C., Liu, L., & Leys, C. (2021). Why Hedges' g_s^* based on the non-pooled standard deviation should be reported with Welch's t-test, *PsyArXiv* https://doi.org/10.31234/osf.io/tu6mp

Fancher, R. (1985). *The intelligence men: makers of the IQ controversy*. W.W. Norton and Company.

Gallardo-Gómez, D., Richardson, R., & Dwan, K. (2024). Standardized mean differences in meta-analysis: a tutorial. *Cochrane Evidence Synthesis and Methods*, *2*, e12047.

Gautschi, W. (1959). Some elementary inequalities relating to the gamma and incomplete gamma function. *Journal of Mathematics and Physics*, *38*, 77–81.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in Social Research*. Sage.

Gómez-Soria, I., Iguacel, I., Cuenca-Zaldívar, J. N., Aguilar-Latorre, A., Peralta-Marrupe, P., Latorre, E., & Calatayud, E. (2023). Cognitive stimulation and psychosocial results in older adults: a systematic review and meta-analysis. *Archives of Gerontology and Geriatrics*, *115*, 105114.

Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: the Cohen's d family. *The Quantitative Methods for Psychology*, *14*, 242–265.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128.

Hedges, L. V. (2025). Interpretation of the standardized mean difference effect size when distributions are not normal or homoscedastic. *Educational and Psychological Measurement*, *85*, 245–257.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Huynh, C. L. (1989). A unified approach to the estimation of effect size in meta-analysis. *Paper presented at the Annual Meeting of the American Educational Research Association*, San Francisco, March 27–31, 1989.

Kaufman, A. S., & Lichtenberger, E. (2006). *Assessing adolescent and adult intelligence*. John Wiley and Sons.

Keselman, H., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, *13*, 110–129.

Kulinskaya, E., & Staudte, R. G. (2007). Confidence intervals for the standardized effect arising in the comparison of two normal populations. *Statistics in Medicine*, *26*, 2853–2871.

Lin, L., & Aloe, A. M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine*, *40*, 403–426.

Noetel, M., Sanders, T., Gallardo-Gómez, D., Taylor, P., del Pozo Cruz, B., Van Den Hoek, D., Smith, J. J., Mahoney, J., Spathis, J., Moresi, M., & Pagano, R. (2024). Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *British Medical Journal*, *384*, e075847.

Peng, C. Y. J., & Chen, L. T. (2014). Beyond Cohen's d: alternative effect size measures for between-subject designs. *The Journal of Experimental Education*, *82*, 22–50.

Satterthwaite, F. (1941). Synthesis of variance. *Psychometrika*, *6*, 309–316.

Shieh, G. (2013). Confidence intervals and sample size calculations for the standardized mean difference effect size between two normal populations under heteroscedasticity. *Behavior Research Methods*, *45*, 955–967.

Suero, M., Botella, J., & Durán, J. I. (2023). Methods for estimating the sampling variance of the standardized mean difference. *Psychological Methods*, *28*, 895–904.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*, 604–634.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, *70*, 129–133.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*, 350–362.

Wilkinson, L., & The Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, *54*, 594–604.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shi, J., Zhang, X., Lin, L., Kwan, H. Y., & Tong, T. (2026). Geometric standardized mean difference and its application to meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *00*, 1–20. https://doi.org/10.1111/bmsp.70071