Annals of the Institute of Statistical Mathematics (2023) 75:281–302
https://doi.org/10.1007/s10463-022-00844-4

Inference using an exact distribution of test statistic for random-effects meta-analysis

Keisuke Hanada[1] · **Tomoyuki Sugimoto**[1]

Received: 31 May 2021 / Revised: 3 June 2022 / Accepted: 8 June 2022 /
Published online: 26 July 2022

Abstract
Random-effects meta-analysis serves to integrate the results of multiple studies with methods such as moment estimation and likelihood estimation duly proposed. These existing methods are based on asymptotic normality with respect to the number of studies. However, the test and interval estimation deviate from the nominal significance level when integrating a small number of studies. Although a method for constructing more conservative intervals has been recently proposed, the exact distribution of test statistic for the overall treatment effect is not well known. In this paper, we provide an almost-exact distribution of the test statistic in random-effects meta-analysis and propose the test and interval estimation using the almost-exact distribution. Simulations demonstrate the accuracy of estimation and application to existing meta-analysis using the method proposed here. With known variance parameters, the estimation performance using the almost-exact distribution always achieves the nominal significance level regardless of the number of studies and heterogeneity. We also propose some methods to construct a conservative interval estimation, even when the variance parameters are unknown, and present their performances via simulation and an application to Alzheimer's disease meta-analysis.

Keywords Exact distribution · Meta-analysis · Random-effects model · Test statistic

1 Introduction

A meta-analysis serves to integrate results from multiple studies to obtain more powerful evidence, for which the fixed- and random-effects models are available. The fixed-effects model assumes that the treatment effects between studies are

✉ Tomoyuki Sugimoto
tomoyuki-sugimoto@biwako.shiga-u.ac.jp

Keisuke Hanada
s7020102@st.shiga-u.ac.jp

[1] Graduate School of Data Science, Shiga University, 1-1-1 Baba, Hikone, Shiga 522-8522, Japan

homogeneous, whereas the random-effects model assumes that there exists some difference in the treatment effects between studies. Although the former is then theoretically simpler and can be used if there is a significant between-study heterogeneity (Whitehead and Whitehead 1991), we should limit it to situations where the assumption of equal treatment effects across studies is valid. For this reason, it is a standard to use the random-effects meta-analysis model, for which several methods exist for estimating the overall treatment effect. For instance, DerSimonian and Laird (1986) estimate the between-study variance based on the weighted mean of treatment effects of each studies, while Paule and Mandel (1982) generalize the moment estimation of overall treatment effect and between-study variance for various weights. Elsewhere, DerSimonian and Kacker (2007) propose estimating the between-study variance through applying two-stage estimation.

All these moment estimation methods insert an estimate of the between-study variance as if it was the true value, whereas the maximum likelihood (ML) method increases its accuracy by simultaneously estimating the overall treatment effect and the between-study variance (Hardy and Thompson 1996). The ML method makes a bias in the estimator of between-study variance, so the restricted ML method ensures the unbiasedness of between-study variance estimator (Thompson and Sharp 1999). Also, the permutation method is often useful when we have a certain number of studies and a need to control Type I error (Follmann and Proschan 1999; Noma et al. 2020).

If the integration of results of many studies is available in meta-analysis, these methods can assume asymptotic normality in the case where the number of studies is sufficiently large. When this does not hold, the coverage probabilities of confidence intervals obtained by these methods fall below the nominal significance level (Michael et al. 2019) and the uncertainty in estimating the between-study heterogeneity increases. Because the best-known DerSimonian-Laird method has a narrow confidence interval when the number of studies is small, Cornell et al. (2014) suggests to use another method, for example Bayesian and Knapp and Hartung (2003) methods.

Figure 1 shows the probability density function (PDF) of test statistics for meta-analysis for various numbers of studies and between-study heterogeneity. With a small number of studies, the standard normal PDF deviates from the simulated PDF, and the t-distribution does not correspond to the case of small heterogeneity among studies. Table 1 supplies the 97.5% points of test statistics. We

Table 1 97.5% points for the test statistic of overall treatment effect in random-effects meta-analysis: the true value calculated by simulation (Simulation), the standard normal distribution (DL), and the t-distribution for $K-1$ degrees of freedom (t_{K-1})

K	I^2	Simulation	DL	t_{K-1}
5	0.3	2.175	1.960	2.776
5	0.9	2.817	1.960	2.776
10	0.3	2.095	1.960	2.262
10	0.9	2.312	1.960	2.262
20	0.3	2.068	1.960	2.093
20	0.9	2.110	1.960	2.093

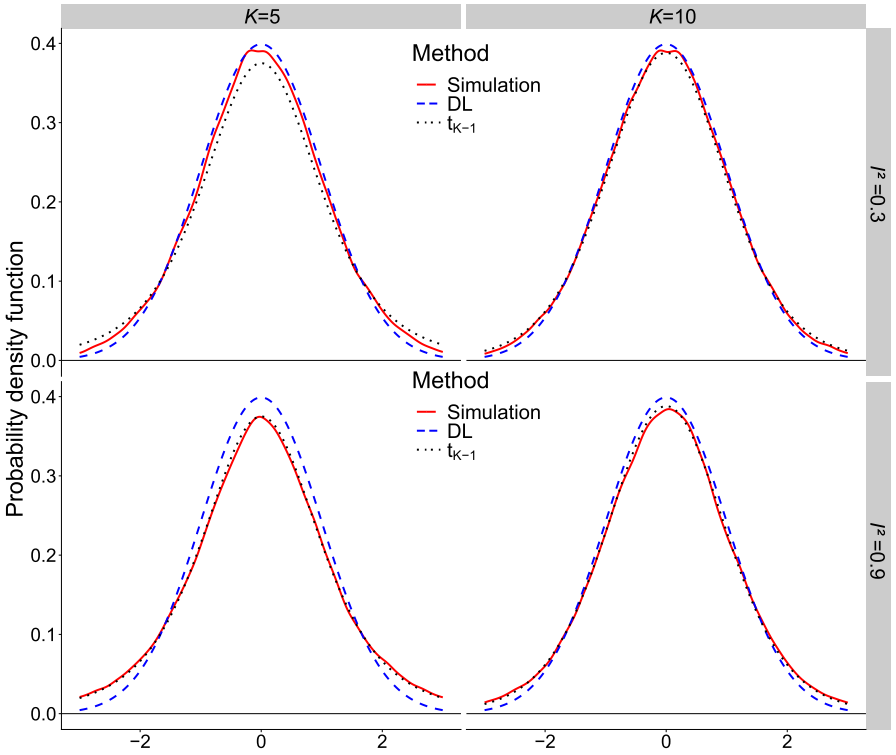

Fig. 1 From left to right, number of tests ($K = 5, 10$), from top to bottom, the simulated histogram (Simulation, solid), the standard normal PDF (DL, dot-dash) of the test statistic for small ($I^2 = 0.3$) and large ($I^2 = 0.9$) heterogeneity, and the PDF of t-distribution with $K - 1$ degrees of freedom (t_{K-1}, dashes)

observe that 97.5% points obtained from the simulation are larger than those for the standard normal distribution, so that the interval estimation and testing using the standard normal distribution is too generous. We find the t-distribution with $K - 1$ degrees of freedom is also inappropriate because it is less than the nominal significance level when the between-study heterogeneity is large. To deal with such cases, it is important to undertake estimation without the use of asymptotic theory.

Michael et al. (2019) propose a conservative confidence interval when the number of studies is small. Hartung (1999) modifies the test statistic to a random variable that follows an approximate t-distribution, allowing for more accurate inference than the moment and ML methods using the normal approximation. Sanchez and Marin (2008) prove the accuracy of confidence intervals for the overall treatment effect using various estimates of the between-study variance by simulation. In summary, when the number of studies is small, the coverage probability of the confidence interval of overall treatment effect deviates from the nominal significance level. This is due to the insertion of between-study variance estimate different from the true value.

In this paper, we present an almost-exact distribution of the test statistic for the DerSimonian-Laird method and compare the accuracy of confidence intervals and p values with existing methods through simulation. We find our approach is always accurate when the true between-study variance parameter as known. We also propose a method conservatively to modify the confidence interval of overall treatment effect by adjusting the naive estimate, even when the parameter is replaced with an estimate of the between-study variance.

The rest of the paper is composed as follows. We introduce the statistical model and reviews existing methods for random-effects meta-analysis in Sect. 2. We present the almost-exact distribution of the between-study variance and the test statistic and suggests methods for test and interval estimation using these exact distributions in Sect. 3. We evaluate the proposed method using simulations and applies it to Alzheimer's disease meta-analysis in Sect. 4, and provide the conclusion in Sect. 5.

2 Model and existing methods

2.1 Model

We focus on the random-effects meta-analysis model available when some between-study heterogeneity exists. Let θ_k denote the true treatment effect in the k-th study ($k = 1, \ldots, K$), and suppose that summary statistics (or observed treatment effect) $\hat{\theta}_k, k = 1, \ldots, K$ are collected for each independent K studies. We can analyze such meta-analysis data using the random-effects model assumed to be as follows:

$$\hat{\theta}_k | \theta_k \sim N(\theta_k, \tau^2), \quad \theta_k \sim N(\theta, \sigma_k^2), \quad k = 1, \ldots, K. \tag{1}$$

where θ is the true value of the overall treatment effect, σ_k^2 is the within-study variance in the k-th study, and τ^2 is the between-study variance. Model (1) can be simply expressed as

$$\hat{\theta}_k \sim N(\theta, \sigma_k^2 + \tau^2).$$

If both τ^2 and σ_k^2, $k = 1, \ldots, K$ are known, the estimator of overall treatment effect $\hat{\theta}$ follows a normal distribution under the null hypothesis $H_0 : \theta = 0$. However, the between-study variance τ^2 is usually unknown in practice, and an exact statistical inference in the case of unknown τ^2 is open in terms of statistical theory and methodology. We discuss the problem and solution for the inference, when τ^2 is unknown, and then present an exact inference procedure for the DerSimonian-Laird method, where it will be assumed that the within-study variances σ_k^2, $k = 1, \ldots, K$ are known for simplicity.

2.2 DerSimonian-Laird method

We briefly explain the best-known DerSimonian-Laird procedure in random-effects meta-analysis. Let $\hat{\theta}_{DL}$ and $\hat{\tau}^2_{DL}$ be the DerSimonian-Laird estimators for true overall treatment effect θ and between-study variance τ^2, which are written as

$$\hat{\theta}_{DL} = \frac{\sum_{k=1}^{K} \frac{\hat{\theta}_k}{\sigma_k^2 + \hat{\tau}^2_{DL}}}{\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \hat{\tau}^2_{DL}}} \quad \text{and} \quad \hat{\tau}^2_{DL} = \max(0, \hat{\tau}^2_{\mathrm{u}}),$$

where $\hat{\tau}^2_{\mathrm{u}}$ is an unbiased estimator for τ^2

$$\hat{\tau}^2_{\mathrm{u}} = \frac{Q - (K-1)}{\gamma_0^{(1)} - \gamma_0^{(2)}/\gamma_0^{(1)}}, \tag{2}$$

Q is a test statistic for between-study heterogeneity,

$$Q = \sum_{k=1}^{K} \frac{(\hat{\theta}_k - \bar{\theta}_0)^2}{\sigma_k^2},$$

$\bar{\theta}_0$ is the inverse-variance weighted average (i.e., $\bar{\theta}_y$ at $y = 0$) using an extend notation

$$\bar{\theta}_y = \textstyle\sum_{k=1}^{K} (\sigma_k^2 + y)^{-1} \hat{\theta}_k / \sum_{k=1}^{K} (\sigma_k^2 + y)^{-1}$$

and $\gamma_0^{(r)}$ is $\gamma_y^{(r)} = \sum_{k=1}^{K} (\sigma_k^{2r} + y)^{-1}$ at $y = 0$ $(r = 1, 2)$. The estimator $\hat{\theta}_{DL}$ is unbiased and has asymptotic efficiency (Jackson et al. 2010), and note that $\hat{\theta}_{DL}$ is $\bar{\theta}_y$ at $y = \hat{\tau}^2_{DL}$ in the extend inverse-variance weighted average.

Let T_{DL} be the test statistic of overall treatment effect for model (1), which is

$$T_{DL} = \frac{\hat{\theta}_{DL}}{\sqrt{1/\gamma^{(1)}_{\hat{\tau}^2_{DL}}}} = \frac{\sum_{k=1}^{K} \frac{\hat{\theta}_k}{\sigma_k^2 + \hat{\tau}^2_{DL}}}{\sqrt{\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \hat{\tau}^2_{DL}}}}.$$

We usually address T_{DL} to be standard normally distributed under the null hypothesis $H_0 : \theta = 0$. Thus, as the $100(1-\alpha)\%$ confidence interval for the overall treatment effect θ, we usually use

$$\hat{\theta}_{DL} \pm z_{\alpha/2} \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \hat{\tau}^2_{DL}} \right)^{-1},$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ point of standard normal distribution. However, the variance of test statistic T_{DL} is 1 if and only if the between-study variance estimator $\hat{\tau}^2_{DL}$ is equal to the true value τ^2. When the number of studies K is small, the asymptotic approximation is not valid yet; therefore, it is not appropriate to use $z_{\alpha/2}$ as

the $100(1-\alpha/2)\%$ point of test statistic. Further, we should note that this procedure does not provide a good estimate for the nominal significance level under a small K.

Advanced methods to improve the accuracy of interval estimation and testing have also been proposed. The two-step method (DerSimonian and Kacker 2007) estimates the between-study variance τ^2 in two steps to improve estimation accuracy. Michael et al. (2019) propose a conservative interval estimation that combines the ML and moment estimation methods. Hartung (1999) suggests modifying the test statistic as approximated by the t-distribution with $K-1$ degrees of freedom. This approximation is better than the normal approximation and can be calculated more accurately, even when the number of studies is small. However, the accuracy is compromised in situations where the within-study variances $\sigma_1^2, \ldots, \sigma_K^2$ substantially differ for each study (Sanchez and Marin 2008).

3 An exact distribution and proposed method

3.1 Exact and almost-exact distributions for the test statistic

We consider an exact distribution of the test statistic T_{DL} obtained without asymptotic approximation which holds true under a large number of studies, and then, we propose an almost-exact version because the exact distribution is difficult to put into practice other than simulation. By the almost-exact version of the exact distribution, we propose a testing and interval estimation procedure to maintain a nominal significance level specified in advance. To start, we provide the exact distribution of the untruncated estimator for between-study variance τ^2.

Lemma 1 *Suppose that the random-effects model (1) is satisfied. The untruncated estimator $\hat{\tau}_{\mathrm{u}}^2$ of between-study variance τ^2 is distributed as*

$$\hat{\tau}_{\mathrm{u}}^2 \sim \frac{\sum_{r=1}^{R} \lambda_r \chi_{r(1)}^2 - (K-1)}{\gamma_0^{(1)} - \gamma_0^{(2)}/\gamma_0^{(1)}},$$

where $\lambda_1, \ldots, \lambda_R$ are nonzero eigenvalues of the matrix VW, V is the variance-covariance matrix of $(\hat{\theta}_1 - \bar{\theta}_0, \ldots, \hat{\theta}_K - \bar{\theta}_0)$, and the (k, k') element of V is

$$V_{kk'} = \begin{cases} \left(\sigma_k^2 - \dfrac{1}{\gamma_0^{(1)}}\right) + \left(1 + \dfrac{\gamma_0^{(2)}}{(\gamma_0^{(1)})^2} - \dfrac{2}{\sigma_k^2 \gamma_0^{(1)}}\right)\tau^2 & (k = k') \\ -\dfrac{1}{\gamma_0^{(1)}} + \left(\dfrac{\gamma_0^{(2)}}{(\gamma_0^{(1)})^2} - \dfrac{1/\sigma_k^2 + 1/\sigma_{k'}^2}{\gamma_0^{(1)}}\right)\tau^2 & (k \neq k'), \end{cases} \tag{3}$$

W is a diagonal matrix with elements $\sigma_1^{-2}, \ldots, \sigma_K^{-2}$, and $\chi_{1(1)}^2, \ldots, \chi_{R(1)}^2$ are mutually independent chi-square random variables with 1 degree of freedom.

This result itself is also considered elsewhere, including Biggerstaff and Jackson (2008). A proof of Lemma 1 is outlined as follows: by applying the theorem of Box (1954) to Q, we can find that the distribution of Q follows $\sum_{r=1}^{R} \lambda_r \chi^2_{r(1)}$. Thus, the distribution of $\hat{\tau}^2_{\mathrm{u}}$ is obtained via simple variable transformation.

Using the exact distribution of $\hat{\tau}^2_{\mathrm{u}}$ in Lemma 1, we provide an expression of the exact distribution of T_{DL} given $\hat{\tau}^2_{\mathrm{u}}$ in Theorem 1. Then, based on Theorem 1, we propose an almost-exact distribution of T_{DL} in Theorem 2. In advance, we prepare a notation

$$\gamma^{(r)}_{y,-k'} = \sum_{\{k:k\neq k'\}} (\sigma_k^{2r} + y)^{-1}$$

required in Theorem 1, which is a quantity except the k'-th element of the sum $\gamma^{(r)}_{y}$.

Theorem 1 *Suppose that the random-effects model* (1) *is satisfied. For mutually independent* $K-1$ *standard normal random numbers* $z_1, \ldots, z_{K-1}$, *let* $T_{1(x)}$ *and* T_2 *be a linear sum statistic and weighted chi-square statistic, respectively, such that*

$$T_{1(x)} = \sum_{k=1}^{K-1} \left\{ \sqrt{\sigma_k^2 + \tau^2}/(\sigma_k^2 + x) \right\} z_k \quad \text{and} \quad T_2 = \sum_{k=1}^{K-1} \left(1 + \tau^2/\sigma_k^2\right) z_k^2.$$

Then, given $\hat{\tau}^2_{\mathrm{u}} = x$, *the test statistic* T_{DL} *is conditionally distributed as*

$$T_{DL}|_{\hat{\tau}^2_{\mathrm{u}}=x} \sim Y(T_{1(0)}, T_{1(x)}; x) + U\varepsilon(T_{1(0)}, T_2; x), \tag{4}$$

where $Y(T_{1(0)}, T_{1(x)}; x)$ *and* $\varepsilon(T_{1(0)}, T_2; x)$ *are functions of* $T_{1(x)}$ *and* T_2 *such that*

$$Y(T_{1(0)}, T_{1(x)}; x) = \left(\theta + \frac{T_{1(x)}}{\gamma^{(1)}_{x,-K}} \right) \frac{\gamma^{(1)}_{x,-K}}{\sqrt{\gamma^{(1)}_{x}}} + \left(\theta + \frac{T_{1(0)}}{\gamma^{(1)}_{0,-K}} \right) \frac{1}{(\sigma_K^2 + x)\sqrt{\gamma^{(1)}_{x}}},$$

$$\begin{aligned} \varepsilon(T_{1(0)}, T_2; x) = \frac{1}{(\sigma_K^2 + x)\sqrt{\gamma^{(1)}_{x}}} \Bigg[& \left(\sigma_K^2 + \frac{1}{\gamma^{(1)}_{0,-K}} \right) \frac{T^2_{1(0)}}{\gamma^{(1)}_{0,-K}} \\ & + 2\theta \left(\sigma_K^2 + \frac{1}{\gamma^{(1)}_{0,-K}} - \sigma_K^2 \gamma^{(1)}_0 \right) T_{1(0)} - \sigma_K^2 \frac{\gamma^{(1)}_0}{\gamma^{(1)}_{0,-K}} T_2 \\ & + \sigma_K^2 \left(\frac{(\gamma^{(1)}_0)^2 - \gamma^{(2)}_0}{\gamma^{(1)}_{0,-K}} \right) x + \sigma_K^2 (K-1) \frac{\gamma^{(1)}_0}{\gamma^{(1)}_{0,-K}} \Bigg]^{\frac{1}{2}} \end{aligned}$$

and U *is a random variable which takes either* 1 *or* -1 *with equal probability* 0.5. *In particular,* $Y(T_{1(0)}, T_{1(x)}; x)$ *is normally distributed with mean* $m_{-K}(x)$ *and variance* $v^2_{-K}(x)$, *where*

$$m_{-K}(x) = \frac{1}{\sqrt{\gamma_x^{(1)}}} \sum_{k=1}^{K-1} \left\{ \frac{1}{\sigma_k^2 + x} + \frac{1}{\sigma_k^2 \gamma_{0,-K}^{(1)}(\sigma_K^2 + x)} \right\} \theta$$

$$v_{-K}^2(x) = \frac{1}{\gamma_x^{(1)}} \sum_{k=1}^{K-1} \left\{ \frac{1}{\sigma_k^2 + x} + \frac{1}{\sigma_k^2 \gamma_{0,-K}^{(1)}(\sigma_K^2 + x)} \right\}^2 (\sigma_k^2 + \tau^2).$$

Remark that $T_{DL}|_{\hat{\tau}_u^2 = x}$ *is not be generated, that is, we have* $f_{T_{DL}|\hat{\tau}_u^2}(t|x) = 0$ *if a value in the square root of* $\varepsilon(T_{1(0)}, T_2; x)$ *is negative, where* $f_{T_{DL}|\hat{\tau}_u^2}(t|x)$ *is the PDF of* $T_{DL}|_{\hat{\tau}_u^2 = x}$.

The proof of Theorem 1 is provided in Appendix 1. From Theorem 1, we observe that the conditional distribution of T_{DL} given $\hat{\tau}_u^2 = x$ is a symmetric but complicated distribution determined by three statistics ($T_{1(0)}$, $T_{1(x)}$, T_2) which follow normal and weighted chi-square distributions. Further, the exact conditional distribution can be constructed by a convolution of one component $Y(T_{1(0)}, T_{1(x)}; x)$ normally distributed and the other random component $U\varepsilon(T_{1(0)}, T_2; x)$. Hence, the exact distribution of T_{DL} implemented using this exact conditional distribution is difficult to put into practice other than simulation. One of reasons that the conditional distribution becomes complicated is because $\hat{\theta}_{DL}$ and $\hat{\tau}_u^2$ are not statistically independent each other. However, the dependence between $\hat{\theta}_{DL}$ and $\hat{\tau}_u^2$ is comparatively mild, and the degree of dependence tends to decrease as K is larger. So, in Theorem 2, we propose an almost-exact distribution of T_{DL} constructed by accepting the normality of the distribution of $T_{DL}|\hat{\tau}_u^2$.

Theorem 2 *Suppose that the random-effects model* (1) *is satisfied. If* K *is not too small* (*for example*, $K \geq 5$), *then the conditional distribution of* T_{DL} *given* $\hat{\tau}_u^2 = x$ *can be approximated by a normal distribution with mean* $\mu(x)$ *and variance* $v^2(x)$, *where*

$$\mu(x) = \theta \sqrt{\gamma_x^{(1)}}, \quad v^2(x) = \frac{1}{\gamma_x^{(1)}} \sum_{k=1}^{K} \sum_{k'=1}^{K} \frac{E[\hat{\theta}_k \hat{\theta}_{k'} | \hat{\tau}_u^2 = x]}{(\sigma_k^2 + x)(\sigma_{k'}^2 + x)}.$$

That is, the PDF of T_{DL}, $f_{T_{DL}}(t; \theta, \tau^2)$ *can be approximated by*

$$\hat{f}_{T_{DL}}(t; \theta, \tau^2) = \int_{-\infty}^{\infty} \phi(t; \mu(x), v^2(x)) f_{\hat{\tau}_u^2}(x) dx, \tag{5}$$

where $\phi(t; \mu(x), v^2(x))$ *is a normal density function with mean* $\mu(x)$ *and variance* $v^2(x)$ *and* $f_{\hat{\tau}_u^2}(x)$ *is the PDF of* $\hat{\tau}_u^2$. *More precisely, as* $K \to \infty$, *the difference between the PDFs* $\hat{f}_{T_{DL}}(t; \theta, \tau^2)$ *and* $f_{T_{DL}}(t; \theta, \tau^2)$ *converges to* 0.

The proof and numerical calculation for Theorem 2 are also provided in Appendices 2 and 3. In particular, the formula and practical calculation of $E[\hat{\theta}_k \hat{\theta}_{k'} | \hat{\tau}_u^2 = x]$ are provided in (8) and (9) of Appendices 2 and 3, respectively.

We compare the conditional distribution of $T_{DL}|\hat{\tau}_u^2 = x$ in Appendix A.4, and hereafter, we focus the statistical inference based on T_{DL}.

3.2 Testing and interval estimation

Using Theorem 2, we can consider the testing and interval estimation procedure for the test statistic T_{DL} to maintain a significance level specified in advance. Here, the null hypothesis is $H_0 : \theta = 0$, and the corresponding alternative hypothesis is $H_1 : \theta > 0$. Even if the overall treatment effect θ is a nonzero case, the zero hypothesis $H_0 : \theta = 0$ is available replacing $\hat{\theta}_k$ with $\hat{\theta}_k - \theta, k = 1, \ldots, K$, where each of the treatment effects is shifted by θ. With Theorem 2, we propose that the one-sided p value is provided by $1 - F_{T_{DL}}(T;\tau^2)$, where T is the actual value obtained for the test statistic. Similarly, the two-sided p value is calculated in the same manner, $2(1 - F_{T_{DL}}(|T|;\tau^2))$. The distribution of test statistic T_{DL} is approximately symmetric under the null hypothesis $H_0 : \theta = 0$ because in the almost-exact PDF we have

$$\begin{aligned}\hat{f}_{T_{DL}}\left(t;0,\tau^2\right) &= \int_{-\infty}^{\infty} \phi(-t;0,v^2(x)) f_{\hat{\tau}_u^2}(x)dx \\ &= \hat{f}_{T_{DL}}\left(-t;0,\tau^2\right).\end{aligned}$$

The $100(1-\alpha)\%$ confidence interval for the overall treatment effect θ is obtained as

$$\hat{\theta}_{DL} \pm t_{DL,\alpha/2}(\tau^2)\sqrt{\sum_{k=1}^{K}(\sigma_k^2 + \hat{\tau}_{DL}^2)^{-1}}, \tag{6}$$

where $t_{DL,\alpha/2}(\tau^2)$ is the $100(1-\alpha/2)\%$ point of test statistic T_{DL} when the between-study variance parameter takes the true value τ^2.

3.3 Handling of the unknown parameter τ^2

We should remark that the true parameter τ^2 is required when determining the critical point, even though we derive an almost-exact confidence interval (6) for θ based on Theorem 2. In actual data analysis, the true between-study variance τ^2 is usually unknown. The simplest solution for addressing this problem is the plug-in approach, which is to insert an estimator $\hat{\tau}_{DL}^2$ of the between-study variance into τ^2. However, the coverage probability of confidence interval is less than the nominal significance level when the between-study variance estimator $\hat{\tau}_{DL}^2$ is inserted. Instead, we consider a more conservative solution to correct the confidence interval. We have a computational difficulty in directly handling the between-study variance τ^2 because the possible range of τ^2 is extensive over $[0, \infty)$. So, we use a heterogeneity measure (Higgins and Thompson 2002)

$$I^2 = \frac{\tau^2}{\sigma_t^2 + \tau^2}$$

rather than directly using the between-study variance τ^2, where σ_t^2 is an averaging index for the within-study variance

$$\sigma_t^2 = (K-1)\gamma_0^{(1)}/(\gamma_0^{(1)2} - \gamma_0^{(2)}).$$

The heterogeneity measure I^2 ranges over [0, 1) and has a one-to-one correspondence with τ^2. Thus, we replace the notation of $t_{DL,\alpha/2}(\tau^2)$ by $t_{DL,\alpha/2}(I^2)$ and denote $100(1-\alpha)\%$ confidence interval (6) by $\mathrm{CI}(I^2)$, respectively, as functions of the heterogeneity measure I^2. We may approximate the $100(1-\alpha/2)\%$ point $t_{DL,\alpha/2}(I^2)$ by

$$t_{DL,\alpha/2}(\hat{I}^2) + t'_{DL,\alpha/2}(\hat{I}^2)(\hat{I}^2 - I^2)$$

using the delta method, where $t'_{DL,\alpha/2}(x)$ is the first derivative of $t'_{DL,\alpha/2}(x)$ w.r.t. x and $\hat{I}^2$ is an estimator of I^2 defined by $\hat{I}^2 = \hat{\tau}^2/(\sigma_t^2 + \hat{\tau}^2)$. However, we find that it is difficult to calculate the expected bias $E[\hat{I}^2 - I^2]$ because the sample distribution of $\hat{I}^2$ is considerably skewed. Further, even if we could estimate the expected bias $E[\hat{I}^2 - I^2]$ better, we had an another unbalanced problem when $\hat{I}^2$ was estimated less than I^2. That is, when estimated $\hat{I}^2$ is smaller than true I^2, the confidence interval constructed by the delta method-based correction fails to include the true parameter I^2 with high frequency, so that the confidence interval could not necessarily achieve the nominal level. So, we consider an approach other than the plug-in method or its delta method-based correction. Prepare a correction value I_c^2 such that $I^2 \leq I_c^2$ as a correction value for $\hat{I}^2$, where the coverage probability $CP(I^2)$ of $100(1-\alpha)\%$ confidence interval for θ is bounded as

$$1 - \alpha = \Pr\{\theta \in CI(I^2)\} \leq \Pr\{\theta \in CI(I_c^2)\}.$$

Hence, we are able to construct a nominal or conservative significance level CI for θ using a correction value I_c^2 instead of true I^2. Next, we propose a construction of I_c^2. Given that the heterogeneity measure I^2 is unknown but follows a prior distribution, we can consider a conditional expectation as a correction value such that

$$I_c^2 = \int_0^1 \max\{I^2, \hat{I}^2\} p(I^2|\hat{I}^2) dI^2 = \frac{\int_0^1 \max\{I^2, \hat{I}^2\} p(I^2) p(\hat{I}^2|I^2) dI^2}{\int_0^1 p(I^2) p(\hat{I}^2|I^2) dI^2},$$

where $p(I^2)$ is a prior distribution of I^2, $p(\hat{I}^2|I^2)$ is the probability of obtaining an estimate $\hat{I}^2$ given the fixed value of true I^2. We choose the standard uniform distribution $U(0, 1)$ as a vague prior distribution for $p(I^2)$ because $I^2 \in [0, 1)$. Nevertheless, we find that a setting of the standard uniform distribution does not always construct a conservative confidence interval. For a more conservative correction, we recommend increasing the lower limit of uniform distribution (e.g.,$U(0.5, 1)$). Using simulation, we provide the results estimated by both prior distributions. The numerical calculation for I_c^2 is based on Monte Carlo integration and including the R program in Supplemental material Sect. 4.

4 Simulation study and application

4.1 Simulation for exact distribution of test statistic

We verify the accuracy of the distribution of test statistic for the overall treatment effect θ using simulation. The 97.5% point of test statistic is given in Table 1. The number of studies and heterogeneity are set from small to large ($K = 5, 10, 20, 30, I^2 = 0.1, 0.3, 0.6, 0.9$). We consistently set the true overall treatment effect θ to 0, which is the target in the test and interval estimation, because the value of θ itself does not affect the accuracy in the inference. We use the within-study variance σ_k^2 as an inconstant setting for each study, $\sigma_k^2 = 1 + 4(k-1)/(K-1), k = 1, \ldots, K$. The number of repetitions in simulation is 100, 000 times.

We can observe the performance of almost-exact PDF (5) from this simulation. Their values match to the first decimal place. The 97.5% point of almost-exact PDF can be calculated accurately, although it is slightly affected by numerical errors. In addition, the standard normal distribution used commonly has a 97.5% point of 1.96 for any study heterogeneity. Therefore, the standard normal distribution provides a 97.5% point different from the exact method, which causes a decrease in accuracy when the number of studies is small. Even when the number of studies is $K = 30$, the 97.5% points of standard normal distribution are smaller than that of the almost-exact PDF. Accordingly, it should be careful to blindly use the standard normal distribution-based method when the number of studies is small. See Supplemental material Sect. 2 for further simulations (Table 2).

4.2 Confidence interval performance when the parameter τ^2 is known

We here verify the accuracy in the confidence interval for the overall treatment effect via simulation. Equation (6) is calculated using the almost-exact distribution of test statistic T_{DL}; thus, it is theoretically accurate. We compare the numerical accuracy of our almost-exact distribution-based method with the DerSimonian-Laird method. Table 3 shows the coverage probabilities calculated by simulation for the 95% confidence interval constructed by three methods based on standard normal distribution, t-distribution, and the almost-exact distribution. The simulation setting is the same as that in Sect. 4.1, but the number of repetitions in simulation is 10, 000 times. The critical point of standard normal distribution-based method may achieve a nominal

Table 2 The 97.5% point of test statistic T_{DL} for the overall treatment effect in the random-effects meta-analysis. 100, 000 simulation-based values (SIM) and the almost-exact distribution (EX) with the proposed eq. (5)

K	$I^2 = 0.1$		$I^2 = 0.3$		$I^2 = 0.6$		$I^2 = 0.9$	
	SIM	EX	SIM	EX	SIM	EX	SIM	EX
5	1.946	1.948	2.178	2.156	2.546	2.525	2.827	2.829
10	1.950	1.925	2.114	2.092	2.255	2.239	2.310	2.312
20	1.942	1.950	2.062	2.042	2.106	2.107	2.122	2.123
30	1.975	1.965	2.045	2.047	2.055	2.035	2.053	2.053

Table 3 The coverage probability of 95% confidence interval of the overall treatment effect. Standard normal distribution (DL), t-distribution with $K-1$ degrees of freedom (t_{K-1}) and almost-exact distribution (EX) with eq. (5)

K	$I^2 = 0.1$			$I^2 = 0.3$			$I^2 = 0.6$			$I^2 = 0.9$		
	DL	t_{K-1}	EX	DL	t_{K-1}	EX	DL	t_{K-1}	EX	DL	t_{K-1}	EX
5	0.955	0.994	0.954	0.926	0.983	0.949	0.887	0.962	0.948	0.875	0.943	0.950
6	0.957	0.990	0.955	0.936	0.981	0.959	0.899	0.957	0.948	0.888	0.946	0.952
7	0.950	0.986	0.949	0.930	0.976	0.952	0.906	0.952	0.947	0.897	0.944	0.949
8	0.950	0.981	0.949	0.928	0.967	0.944	0.909	0.950	0.947	0.903	0.946	0.950
9	0.951	0.978	0.947	0.931	0.967	0.950	0.914	0.951	0.946	0.910	0.948	0.952
10	0.957	0.979	0.953	0.934	0.966	0.956	0.916	0.950	0.952	0.915	0.947	0.950

confidence level when the heterogeneity is small, but the actual coverage probability decreases less than for the nominal level as heterogeneity increases. Further, the critical point of t-distribution-based method with $K-1$ degrees of freedom is more conservative than that of standard normal distribution-based method but becomes too conservative when the heterogeneity is small. For our almost-exact distribution-based method, the coverage probability is at the nominal significance level regardless of the number of studies and the study heterogeneity, even when affected by numerical errors. We remark again that the almost-exact distribution includes the between-study variance parameter.

4.3 The confidence interval when the parameter τ^2 is unknown

We now verify the accuracy of confidence interval for the overall treatment effect when the parameter τ^2 is unknown. We construct three confidence intervals for the overall treatment effect based on the almost-exact distribution as follows: CI(I^2) which uses the true heterogeneity measure I^2, CI($\hat{I}^2$) which uses the estimated value $\hat{I}^2$ instead of the true I^2, and CI(I_c^2) constructed by the conservative correction I_c^2. We compare the accuracy of three approaches using the coverage probabilities. Among the existing methods, we also compare the DerSimonian-Laird method (DL), the restricted ML method (REML), the Michael et al. (2019) method, which implements conservative inference, and the Hartung (1999) method, which corrects the test statistic for the overall treatment effect (HKSJ).

The simulation result for the coverage probability and the length of 95% confidence interval is provided in Fig. 2. We compare the performances of CI(I^2) constructed using the true I^2 (EX(I2)), CI($\hat{I}^2$) where I^2 is replaced by the estimated value $\hat{I}^2$ (EX(I2hat)), CI(I_c^2) using conservative correction values, and the existing methods. The conservative correction value for CI(I_c^2) depends on the prior distribution. We show the case of a vague prior distribution $U(0, 1)$ (EX(I2c1)) and more restricted prior $U(0.5, 1)$ (EX(I2c2)). The simulation setup is the same as those in Sect. 4.2.

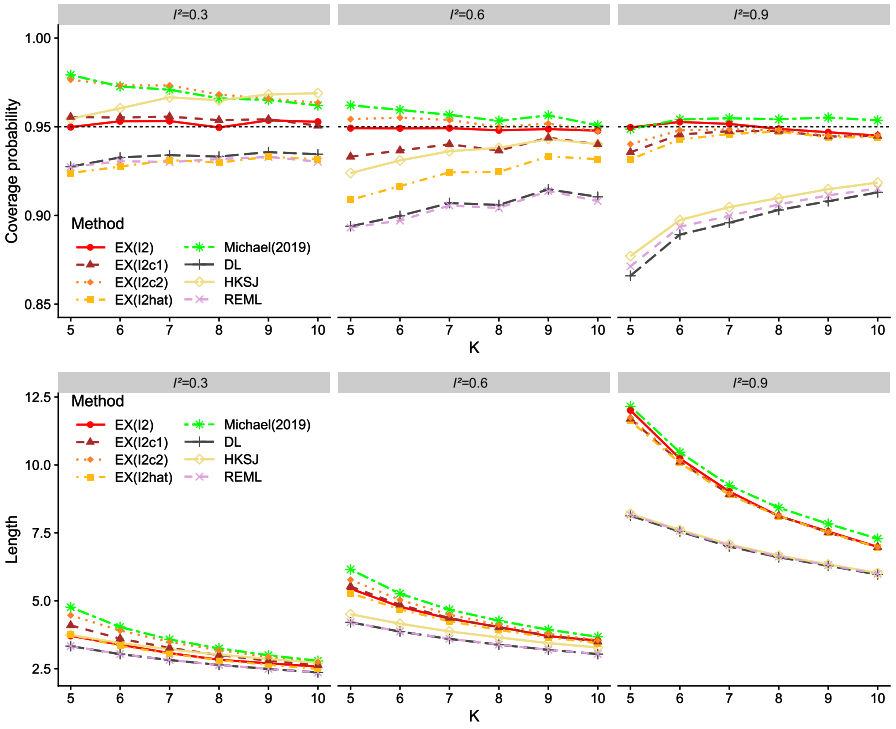

Fig. 2 Simulation results for the coverage probability of 95% confidence interval (top) and the length of confidence interval (bottom). Heterogeneity is small (($I^2 = 0.3$, left), medium ($I^2 = 0.6$, center), and large ($I^2 = 0.9$, right)

EX(I2) calculates the coverage probability at the nominal significance level regardless of the number of studies and the level of heterogeneity. EX(I2hat) provides the coverage probability at the nominal significance level when the heterogeneity is large. As shown, the estimation accuracy of EX(I2hat) is the same as or better than the DerSimonian-Laird method. Moreover, EX(I2c1) has the estimation performance as well as the HKSJ method when the heterogeneity is moderate and can estimate at the nominal significance level when the heterogeneity is small or large. EX(I2c2) is more conservative in the proposed methods, consistently constructing confidence intervals that are either at the nominal or conservative significance level.

Elsewhere, the Michael et al. method results in too conservative confidence intervals when the heterogeneity is small, although it is useful to keep the coverage probability above the nominal significance level for any heterogeneity. Overall, if we wish to consistently maintain some conservative inference, we can recommend EX(I2c2) or the Michael et al. methods as better selectable options. See Supplemental material Sects. 3 and 4 for R program code of the proposed method and further simulations.

4.4 Application

Including our proposed method, we provide an application to an existing meta-analysis data in Chen et al. (2017). Their meta-analysis highlights 11 study results for Alzheimer's disease, which evaluate the treatment effect using Hedge's g that is an index used when comparing two samples,

$$g = \frac{|\bar{x}_1 - \bar{x}_0|}{SD_g},$$

where the treatment group average effect is $\bar{x}_1$, the control group average effect is $\bar{x}_0$, and we calculate

$$SD_g = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_0 - 1)SD_0^2}{n_1 + n_0 - 2}}$$

using the sample sizes n_1, n_0 and estimated standard deviations SD_1, SD_0 of treatment and control groups, respectively.

We apply the proposed methods to this meta-analysis data and compare the results for REML, DL, HKSJ, the Michael et al. method, and the proposed method (EX(I2hat), EX(I2c1), EX(I2c2)). Table 4 presents the results for 95% confidence interval. Also, we applied the REML, DL, and HKSJ methods using the package "metafor" in the programming language R for statistical analysis, and the Michael et al. method using the package "meta.exact".

As shown in Table 4, the REML and DL methods yield about the same confidence intervals, while those for the Michael et al. and proposed methods construct more conservative intervals. The HKSJ method also provides a conservative confidence interval and the p value is smaller than the proposed methods. Considering the numerical experiments, we have to pay attention to the traditional methods construct narrower confidence intervals than the almost-exact distribution-based method.

Table 4 Results of applying each 95% confidence interval to meta-analysis of Chen et al. (2017) for Alzheimer's disease

Method	Mean	SD	95% CI		Test statistic	p value
DL	0.413	0.108	0.201	0.625	3.813	0.00014
REML	0.422	0.121	0.185	0.658	3.490	0.00048
HKSJ	0.441	0.159	0.130	0.751	2.780	0.00544
Michael et al. (2019)			0.152	0.682		
EX(I2hat)	0.413	0.108	0.150	0.676	3.813	0.00803
EX(I2c1)	0.413	0.108	0.155	0.671	3.813	0.00663
EX(I2c2)	0.413	0.108	0.154	0.672	3.813	0.00648

5 Conclusion

We construct the exact and almost-exact distribution of test statistic for the DerSimonian-Laird method as one of the moment estimation methods and modify it to compute the test and confidence interval more accurately. The almost-exact distribution of test statistic is calculated from the distribution of conditional test statistic, given an estimate and the true value of between-study variance.

Inference based on the almost-exact distribution of test statistic has the limitation that it requires the between-study variance parameter to be possibly unknown. If we can use the true value of between-study variance, the testing and interval estimation can be consistently performed at the nominal significance level. In contrast, when we use an estimate for the between-study variance, the coverage probability of confidence interval deviates from the nominal significance level, especially when there is large between-study heterogeneity and the number of studies is small. We propose the use of conditional expectation of the heterogeneity measure as an index parameter to characterize the almost-exact distribution of test statistic. We show that a conservative confidence interval for the overall treatment effect can be constructed by the proposed index parameter. A conservative inference for random-effects meta-analysis has also been proposed by Michael et al. (2019) and Rover et al. (2015).

Our proposed method is supported by two main assumptions. One is that the treatment effects for each of studies in the meta-analysis follow a normal distribution, another is that the within-study variances $(\sigma_1^2, \ldots, \sigma_K^2)$ are known. The assumption of normality is preferably used if the treatment effect is the mean of continuous observations, but this is not available directly for binary or survival data. For noncontinuous data, the assumption of normality may be recovered via some transformation, such as the log-odds ratio for binary data (Stijnen 2010).

As another problem, we include the assumption that the within-study variances $(\sigma_1^2, \ldots, \sigma_K^2)$ are known. Although this is acceptable if the sample size in each study is sufficiently large, the assumption is violated if all the sample sizes are small. For such a solution, we need to calculate the other version of almost-exact distribution when the true values of within-study variances are replaced by their estimates. However, the integration of exact distribution for each within-study variance would require an integral calculation for each and the computation time would be many times longer than if only the between-study variance is assumed to be unknown.

Although the proposed method can be applied even when these two assumptions do not hold, it should be noted that the accuracy of method cannot be ensured. If the model (1) assumptions are inappropriate due to publication bias, it is necessary to select a more appropriate method (Li et al. 2022). Another important issue is that the between-study variance τ^2 is included as a parameter to determine the exact distribution of test statistic. If we could construct a method which does not use the variance parameter, such as for Student's t test, then we would be able to perform the meta-analysis more accurately regardless of the between-study heterogeneity. The method of approximating t-distribution of

Hartung (1999) is based on this idea, but the coverage probability varies depending on the heterogeneity. It would then be useful to have a nominal significance level of the coverage probability irrespective of the between-study heterogeneity, and it would be an interesting future research question.

6 Supplementary material

Supplementary material available at online includes more simulation results and simulation program of R.

A Theoretical details and numerical calculation

A.1 Proof of Theorem 1

First, we should note that $\hat{\theta}_{DL}$ and $\hat{\tau}^2_{\mathrm{u}}$ are not mutually independent. This situation is quite complicated, unlike the case of sample mean and unbiased variance from normal data with equal error-variances, such as Student's t test. For a such reason, in order to derive the distribution of T_{DL}, we consider the marginalization

$$f_{T_{DL}}(t) = \int_{-\infty}^{\infty} f_{T_{DL}|\hat{\tau}^2_{\mathrm{u}}}(t|x) f_{\hat{\tau}^2_{\mathrm{u}}}(x) dx$$

by decomposing the joint PDF of $(T_{DL}, \hat{\tau}^2_{\mathrm{u}})$. Next, we discuss the mean and variance of $T_{DL}|\hat{\tau}^2_{\mathrm{u}}$ (i.e., the conditional random variable of T_{DL} given $\hat{\tau}^2_{\mathrm{u}}$). Given $\hat{\tau}^2_{\mathrm{u}} = x$, we have the constraint condition

$$\sum_{k=1}^{K} \frac{(\hat{\theta}_k - \bar{\theta}_0)^2}{\sigma_k^2} = \left(\gamma_0^{(1)} - \frac{\gamma_0^{(2)}}{\gamma_0^{(1)}} \right) x + (K-1) \tag{7}$$

imposed on $\hat{\theta}_k$, $k = 1, \ldots, K$. By solving (7) on $\hat{\theta}_{k'}$ $(k' = 1, \ldots, K)$, we obtain two cases

$$\hat{\theta}_{k'} = \bar{\theta}_{0,-k'} \pm \xi_{k'}(\bar{\theta}_{0,-k'}, v^2_{-k'}; x)$$

where $\bar{\theta}_{y,-k'}$ and $v^2_{-k'}$ are an extended inverse-variance weighted average and its square average version except the k'-th element $\hat{\theta}_{k'}$, respectively, that is,

$$\bar{\theta}_{y,-k'} = \frac{\sum_{\{k:k\neq k'\}} (\sigma_k^2 + y)^{-1} \hat{\theta}_k}{\sum_{\{k:k\neq k'\}} (\sigma_k^2 + y)^{-1}} \quad \text{and} \quad v^2_{-k'} = \frac{\sum_{\{k:k\neq k'\}} \sigma_k^{-2} \hat{\theta}_k^2}{\sum_{\{k:k\neq k'\}} \sigma_k^{-2}},$$

$\xi_{k'}(\bar{\theta}_{0,-k'}, v^2_{-k'}; x)$ is a quantity written as

$$\xi_{k'}(\bar{\theta}_{0,-k'}, v^2_{-k'};x) = \sqrt{\bar{\theta}^2_{0,-k'} + \sigma^2_{k'}\gamma^{(1)}_{0,-k'}\left\{\bar{\theta}^2_{0,-k'} - \frac{\gamma^{(1)}_0}{\gamma^{(1)}_{0,-k'}}v^2_{-k'} + \left(\frac{(\gamma^{(1)}_0)^2 - \gamma^{(2)}_0}{(\gamma^{(1)}_{0,-k'})^2}\right)x + (K-1)\frac{\gamma^{(1)}_0}{(\gamma^{(1)}_{0,-k'})^2}\right\}}.$$

See Supplemental material Sect. 5 for further details of solving (7).

For simplicity, let $k' = K$. The PDF of $\hat{\theta}_{k'}|\hat{\tau}^2_u$ can be expressed as

$$\begin{aligned}
&f_{\hat{\theta}_K|\hat{\tau}^2_u}(z_K - \theta|x)\\
&= \int\cdots\int_{\Omega} f_{\hat{\theta}_1,\cdots,\hat{\theta}_{K-2},\hat{\tau}^2_u,\hat{\theta}_K}(z_1-\theta,\ldots,z_{K-2}-\theta,x,z_K-\theta)dz_1\cdots dz_{K-2}/f_{\hat{\tau}^2_u}(x)\\
&= \frac{\sigma^2_{K-1}((\gamma^{(1)}_0)^2 - \gamma^{(2)}_0)x}{f_{\hat{\tau}^2_u}(x)}\int\cdots\int_{\Omega}\left(\prod_{k=1}^{K-2}\phi_k(z_k-\theta)\right)\phi_K(z_K-\theta)\\
&\times\left\{\phi_{K-1}(\bar{z}_{-(K-1)} + \xi_{-(K-1)}(\mathbf{z})) + \phi_{K-1}(\bar{z}_{-(K-1)} - \xi_{-(K-1)}(\mathbf{z}))\right\}\xi^{-1}_{-(K-1)}(\mathbf{z})dz_1\cdots dz_{K-2}\\
&= \frac{\sigma^2_{K-1}((\gamma^{(1)}_0)^2 - \gamma^{(2)}_0)x}{f_{\hat{\tau}^2_u}(x)}\int\cdots\int_{\Omega}\left(\prod_{k=1}^{K-2}\phi_k(-z_k-\theta)\right)\phi_K(-z_K-\theta)\\
&\times\left\{\phi_{K-1}(\bar{z}_{-(K-1)} + \xi_{-(K-1)}(\mathbf{z})) + \phi_{K-1}(\bar{z}_{-(K-1)} - \xi_{-(K-1)}(\mathbf{z}))\right\}\xi^{-1}_{-(K-1)}(\mathbf{z})dz_1\cdots dz_{K-2}\\
&= f_{\hat{\theta}_K|\hat{\tau}^2_u}(-z_K-\theta|x),
\end{aligned}$$

where $\Omega = \{(\hat{\theta}_1,\ldots,\hat{\theta}_K)\in\mathbb{R}^n|\hat{\tau}^2_u = x\}$, $\bar{z}_{-(K-1)} = \bar{\theta}_{0,-(K-1)}(z_1,\ldots,z_{K-2},z_K)$, $\xi_{-(K-1)}(\mathbf{z}) = \xi_{-(K-1)}(z_1,\ldots,z_{K-2},x,z_K)$ and ϕ_k is the PDF of $N(\theta,\sigma^2_k+\tau^2)$. See Supplemental material Sect. 6 for further derivation details of $f_{\hat{\theta}_K|\hat{\tau}^2_u}$. We have $E[\hat{\theta}_K|\hat{\tau}^2_u] = \theta$ because the distribution of $\hat{\theta}_K|\hat{\tau}^2_u$ is symmetry. That is, two cases of $\hat{\theta}_K$ obtained under the constraint condition (7) must occur fairly in probabilistic events. Hence, we observe that $T_{DL}|\hat{\tau}^2_u$ can be transformed to

$$T_{DL}|_{\hat{\tau}^2_u=x} = Y(\bar{\theta}_{x,-k'},\bar{\theta}_{0,-k'};x) + U\varepsilon(\bar{\theta}_{0,-k'},v^2_{-k'};x)$$

using U which takes either 1 or -1 with equal probability, where

$$\begin{aligned}
Y(\bar{\theta}_{x,-k'},\bar{\theta}_{0,-k'};x) &= \bar{\theta}_{x,-k'}\frac{\gamma^{(1)}_{x,-k'}}{\sqrt{\gamma^{(1)}_x}} + \frac{\bar{\theta}_{0,-k'}}{(\sigma^2_{k'}+x)}\frac{1}{\sqrt{\gamma^{(1)}_x}},\\
\varepsilon(\bar{\theta}_{0,-k'},v^2_{-k'};x) &= \frac{\xi_{-k'}(x,\bar{\theta}_{0,-k'},v^2_{-k'})}{(\sigma^2_{k'}+x)}\frac{1}{\sqrt{\gamma^{(1)}_x}}.
\end{aligned}$$

The distribution of $\varepsilon(\bar{\theta}_{0,-k'},v^2_{-k'};x)$ is determined by two quantities $\bar{\theta}_{0,-k'}$ and $v^2_{-k'}$ but cannot be expressed by some known distribution. On the other hand, because $Y(\bar{\theta}_{x,-k'},\bar{\theta}_{0,-k'};x)$ is composed of a linear sum of $\bar{\theta}_{x,-k'}$ and $\bar{\theta}_{0,-k'}$, $Y(\bar{\theta}_{x,-k'},\bar{\theta}_{0,-k'};x)$ is normally distributed and we find that it has mean $m_{-k'}(x)$ and variance $v^2_{-k'}(x)$. Replacing $\hat{\theta}_k$ with $\hat{\theta}_k = \theta + \sqrt{\sigma^2_k+\tau^2}z_k$ using mutually independent random numbers $z_1,\cdots,z_K\sim N(0,1)$, three quantities $\bar{\theta}_{x,-k'}$, $\bar{\theta}_{0,-k'}$ and $v^2_{-k'}$ to determine the distribution of $T_{DL}|_{\hat{\tau}^2_u=x}$ are written as

$$\bar{\theta}_{x,-k'} = \theta + \gamma_x^{(1)^{-1}} \sum_{\{k:k\neq k'\}} z_k \sqrt{\sigma_k^2+\tau^2}/(\sigma_k^2+x)$$
$$v^2_{-k'} = (\gamma^{(1)}_{0,-k'})^{-1}\left[\theta^2\gamma^{(1)}_{0,-k'} + 2\theta \sum_{\{k:k\neq k'\}} z_k\sqrt{\sigma_k^2+\tau^2}/\sigma_k^2 + \sum_{\{k:k\neq k'\}} z_k^2(\sigma_k^2+\tau^2)/\sigma_k^2\right].$$

We obtain the result of this theorem by replacing $\bar{\theta}_{0,-k'}$ and $v^2_{-k'}$ in $\xi_{k'}(\bar{\theta}_{0,-k'}, v^2_{-k'};x)$ with

$$\bar{\theta}_{0,-k'} = \frac{T_{1(0)}}{\gamma^{(1)}_{0,-k'}} + \theta \quad \text{and} \quad v^2_{-k'} = \frac{T_2}{\gamma^{(1)}_{0,-k'}} + 2\theta T_{1(0)} + \theta^2$$

for $k' = K$ and using the relation $\gamma_0^{(1)} - \gamma^{(1)}_{0,-K} = \sigma_K^{-2}$. □

A.2 Proof of Theorem 2

An idea of Theorem 2 is to approximate the distribution of $T_{DL}|_{\hat{\tau}_u^2=x}$ by a normal distribution, because the exact conditional distribution is a complicated but symmetric distribution. Concretely, the mean and variance of $T_{DL}|_{\hat{\tau}_u^2=x}$ are, respectively, obtained as

$$\mu(x) = E[T_{DL}|\hat{\tau}_u^2 = x] = \frac{\sum_{k=1}^K \frac{E[\hat{\theta}_k|\hat{\tau}_u^2=x]}{\sigma_k^2+x}}{\sqrt{\sum_{k=1}^K \frac{1}{\sigma_k^2+x}}} = \theta\sqrt{\sum_{k=1}^K \frac{1}{\sigma_k^2+x}},$$

and

$$v^2(x) = V[T_{DL}|\hat{\tau}_u^2 = x] = \frac{1}{\sum_{j=1}^K(\sigma_j^2+x)^{-1}} \sum_{k=1}^K\sum_{k'=1}^K \frac{E[\hat{\theta}_k\hat{\theta}_{k'}|\hat{\tau}_u^2=x]}{(\sigma_k^2+x)(\sigma_{k'}^2+x)},$$

where the form of $E[\hat{\theta}_i\hat{\theta}_j|\hat{\tau}_u^2 = x]$ is

$$E[\hat{\theta}_i\hat{\theta}_j|\hat{\tau}_u^2 = x] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} t_i t_j \frac{f_{(\hat{\theta}_1,\ldots,\hat{\theta}_{K-1},\hat{\tau}^2)}(t_1,\ldots,t_{K-1},x)}{f_{\hat{\tau}^2}(x)} dt_1\cdots dt_{K-1}, \quad (8)$$

and the joint PDF of $(\hat{\theta}_1,\ldots,\hat{\theta}_{K-1},\hat{\tau}_u^2)$ is written as

$$\begin{aligned} &f_{(\hat{\theta}_1,\ldots,\hat{\theta}_{K-1},\hat{\tau}_u^2)}(t_1,\ldots,t_{K-1},x) \\ &= \frac{\sigma_K^2(\gamma_0^{(1)2} - \gamma_0^{(2)})}{4\gamma^{(1)}_{0,-K}\xi_K(\bar{t}_{-K},\overline{t^2}_{-K};x)} \prod_{k=1}^{K-1}\phi\left(t_k;\theta,\sigma_k^2+\tau^2\right) \\ &\quad\times\left\{\phi\left(\bar{t}_{-K}+\xi_K(\bar{t}_{-K},\overline{t^2}_{-K};x);\theta,\sigma_K^2+\tau^2\right)\right. \\ &\quad\left.+\phi\left(\bar{t}_{-K}-\xi_K(\bar{t}_{-K},\overline{t^2}_{-K};x);\theta,\sigma_K^2+\tau^2\right)\right\}, \end{aligned}$$

where $\bar{t}_{-K} = \left(\sum_{k=1}^{K-1} t_k/\sigma_k^2\right)/\gamma_{0,-K}^{(1)}$ and $\overline{t^2}_{-K} = \left(\sum_{k=1}^{K-1} t_k^2/\sigma_k^2\right)/\gamma_{0,-K}^{(1)}$ Thus, the conditional mean $\mu(x)$ and variance $\nu^2(x)$ in Theorem 2 are obtained in this manner.

Finally, we show that the distribution of $T_{DL}|\hat{\tau}_{\mathrm{u}}^2$ can be approximated by normal distribution. For simplicity, assume $\sigma_1^2 \leq \cdots \leq \sigma_K^2$, so that we have $\sum_k^2(\sigma_k^2 + x)^{-1} \geq K(\sigma_K^2 + x)^{-1}$. Denote a sum of normally distributed random variables as an approximate value of $T_{DL}|_{\hat{\tau}_{\mathrm{u}}^2=x}$ by

$$\tilde{T}_{DL}|_{\hat{\tau}_{\mathrm{u}}^2=x} = \frac{\sum_{k=1}^{K-1} \hat{\theta}_k/(\sigma_k^2 + x)}{\sqrt{\sum_{k=1}^{K} 1/(\sigma_k^2 + x)}}.$$

Then, we have

$$\begin{aligned}\left|T_{DL}|_{\hat{\tau}_{\mathrm{u}}^2=x} - \tilde{T}_{DL}|_{\hat{\tau}_{\mathrm{u}}^2=x}\right| &= \left|\frac{\{\bar{\theta}_{0,-K} \pm \xi_K(\bar{\theta}_{0,-K}, \nu_{-K}^2; x)\}/(\sigma_K^2 + x)}{\sqrt{\sum_{k=1}^{K} 1/(\sigma_k^2 + x)}}\right| \\ &\leq |cK^{-1/2}| \to 0 \quad (K \to \infty),\end{aligned}$$

where c is a constant. In Euclidean distance, $T_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}$ converges to $\tilde{T}_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}$ (as $K \to \infty$). This means that $\lim_{K\to\infty} |f_{T_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}}(t|x) - f_{\tilde{T}_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}}(t|x)| = 0$ in the PDFs of two random variables by the continuity of the function $f_{T_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}}(t|x)$. Also, it is clear that $f_{\tilde{T}_{DL}|_{\hat{\tau}_{\mathrm{u}}^2}}(t|x)$ and $\phi(t;\mu(x), \nu^2(x))$ are equivalent for sufficiently large K by showing that they have identical mean and variance. Therefore, the proof of Theorem 2 is complete. □

A.3 Numerical calculation for Theorem 2

We have two problems in numerical calculation for Theorem 2. One is to calculate the PDF of between-study variance $\hat{\tau}_{\mathrm{u}}^2$, and another is to calculate the variance of $T_{DL}|\hat{\tau}_{\mathrm{u}}^2$. The estimate of between-study variance, $\hat{\tau}_{\mathrm{u}}^2$, is the sum of weighted chi-square distribution, as shown by Lemma 1. The numerical calculation of the PDF of weighted chi-square distribution is proposed by many authors, for example, Imhof (1961), Akkouchi (2005), and so on. However, these methods are not necessarily accurate when the number of studies K is small. Therefore, we calculate the PDF by generating random numbers that follow the chi-square distribution with 1 degree of freedom using simulation. That is, if $x_k(\ell), k = 1, \ldots, K, \ell = 1, \ldots, L$ are random numbers that independently follow a chi-square distribution with 1 degree of freedom, then the random number $y_1, \ldots, y_L$ of L estimates of the between-study variance represents

$$y_\ell = \frac{\sum_{r=1}^{R} \lambda_r x_r(\ell) - (K-1)}{\gamma_0^{(1)} - \gamma_0^{(2)}/\gamma_0^{(1)}}.$$

Thus, the cumulative distribution function of the between-study variance estimate $\hat{\tau}_{\mathrm{u}}^2$ is

$$F_{\hat{\tau}^2_{\mathrm{u}}}(x) = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}\{y_\ell \le x\}.$$

The PDF $f_{\hat{\tau}^2_{\mathrm{u}}}$ is also calculated using the numerical differentiation, $f_{\hat{\tau}^2_{\mathrm{u}}}(x) = (F_{\hat{\tau}^2_{\mathrm{u}}}(x+h) - F_{\hat{\tau}^2_{\mathrm{u}}}(x))/h$, where h is a sufficiently small constant. In our simulation, we set $L = 100,000$ for calculating $f_{\hat{\tau}^2_{\mathrm{u}}}$.

In the numerical calculation of the variance of $T_{DL}|\hat{\tau}^2_{\mathrm{u}}$, Monte Carlo integration with importance sampling is quite useful and a handy tool in this case. That is, if $z_k^{(\ell)}, \ell = 1, \ldots, L, k = 1, \ldots, K-1$ are random numbers sampled from standard normal distribution $N(0, 1)$, the mean of $\hat{\theta}_i\hat{\theta}_j|\hat{\tau}^2_{\mathrm{u}}$ can be approximated as

$$\begin{aligned} &E[\hat{\theta}_i\hat{\theta}_j|\hat{\tau}^2_{\mathrm{u}} = x] \\ &\doteq \frac{1}{L}\sum_{\ell=1}^{L} \frac{\sigma_K^2((\gamma_0^{(1)})^2 - \gamma_0^{(2)})}{4\gamma^{(1)}_{0,-K}\xi_K(\bar{z}^{(\ell)}_{-K}, \overline{z^2}^{(\ell)}_{-K};x)} \prod_{k=1}^{K-1} \phi\left(\theta + \sqrt{\sigma_k^2+\tau^2}z_k^{(\ell)};\theta, \sigma_k^2+\tau^2\right) \\ &\times \left\{\phi\left(\bar{z}^{(\ell)}_{-K} + \xi_K(\bar{z}^{(\ell)}_{-K}, \overline{z^2}^{(\ell)}_{-K};x);\theta, \sigma_K^2+\tau^2\right)\right. \\ &\left.+\phi\left(\bar{z}^{(\ell)}_{-K} - \xi_K(\bar{z}^{(\ell)}_{-K}, \overline{z^2}^{(\ell)}_{-K};x);\theta, \sigma_K^2+\tau^2\right)\right\}, \end{aligned} \tag{9}$$

where

$$\begin{aligned} \bar{z}^{(\ell)}_{-K} &= \theta + \gamma^{(1)-1}_{0,-K}\left(\sum_{k=1}^{K-1}\sqrt{\sigma_k^2+\tau^2}z_k^{(\ell)}/\sigma_k^2\right) \\ \overline{z^2}^{(\ell)}_{-K} &= \gamma^{(1)-1}_{0,-K}\left(\sum_{k=1}^{K-1}(\theta + \sqrt{\sigma_k^2+\tau^2}z_k^{(\ell)})^2/\sigma_k^2\right). \end{aligned}$$

Thus, the conditional variance $v^2(x)$ is calculated based on (9) given sufficiently large L. In our simulation, we set $L = 10,000$ for calculating $E[\hat{\theta}_i\hat{\theta}_j|\hat{\tau}^2_{\mathrm{u}} = x]$. Using the two PDFs $f_{\hat{\tau}^2_{\mathrm{u}}}$ and $f_{T_{DL}|\hat{\tau}^2_{\mathrm{u}}}$, the PDF (5) of test statistic T_{DL} can be calculated, where the trapezoidal or Simpson's rules are useful for numerical integration using these PDFs.

A.4 Simulation of conditional test statistic $T_{DL}|\hat{\tau}^2_u = x$

We compare the difference between the exact conditional distribution of $T_{DL}|\hat{\tau}^2_u = x$ and its normal approximation under some x. The setting of θ and σ_k^2 are the same as Sect. 4.2. The number of study is set to small case $K = 5$ and the heterogeneity is set to small and large $I^2 = 0.3, 0.9$. x is set to (1, 2) for comparing various situations. We sample 100,000 data of $(z_1, \ldots, z_{k-1})$ from standard normal distribution and calculate the exact distribution of $T_{DL}|\hat{\tau}^2_u = x$. The almost-exact distribution of $T_{DL}|\hat{\tau}^2_u = x$ can be calculated by the method in Appendix A.3.

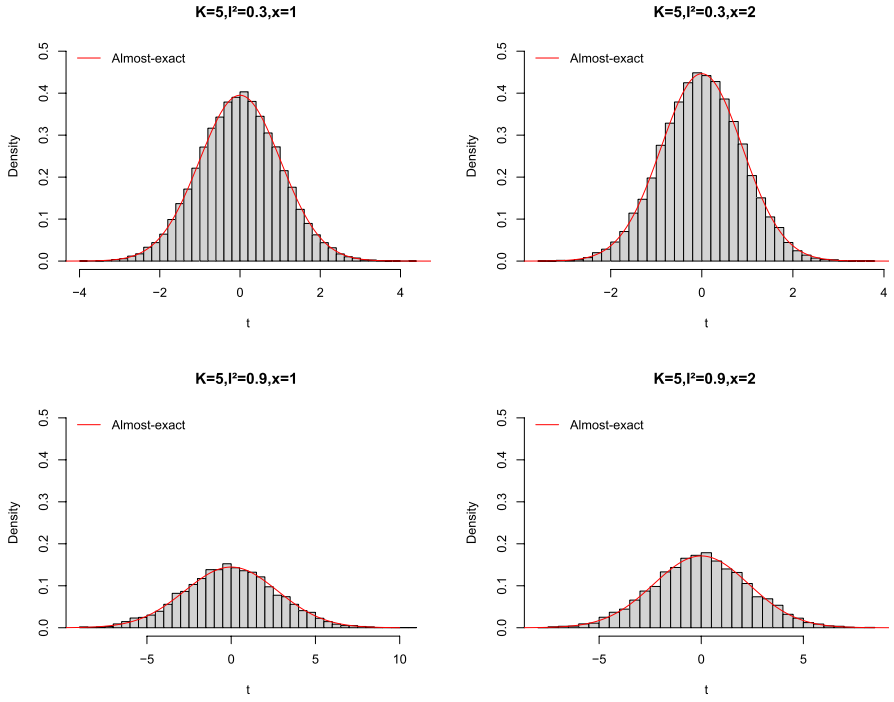

Fig. 3 Simulation histogram of conditional test statistic $T_{DL}|\hat{\tau}_u^2 = x$ and the almost-exact PDF (red line)

We show the simulation histogram based on the exact conditional distribution and the corresponding almost-exact PDF in Fig. 3. The almost-exact distribution of $T_{DL}|\hat{\tau}_u^2 = x$ approximates its exact distribution well in the small number of studies. This provides that it is reasonable to use the almost-exact distribution of $T_{DL}|\hat{\tau}_u^2 = x$ for our simulations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00844-4.

Acknowledgements We thank the editor, the associate editor and the two anonymous referees for their constructive comments. This work was supported by JSPS KAKENHI (Grant number JP21K11783).

References

Akkouchi, M. (2005). On the convolution of gamma distributions. *Soochow Journal of Mathematics, 31*, 205–211.

Biggerstaff, B., Jackson, D. (2008). The exact distribution of Cochran's heterogeneity statistic in one-way random effects meta-analysis. *Statistics in Medicine, 27*, 6093–6110.

Box, G. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics, 25*, 290–302.

Chen, R., Chan, P., Chu, H. (2017). Treatment effects between monotherapy of donepezil versus combination with memantine for alzheimer disease: A meta-analysis. *PLOS ONE, 12*, 1–14.

Cornell, J., Mulrow, C., Localio, R. (2014). Random-effects meta-analysis of inconsistent effects: A time for change. *Annals of Internal Medicine, 160*, 267–270.

DerSimonian, R., Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials, 28*, 105–114.

DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177–188.

Follmann, D., Proschan, M. (1999). Valid inference in random effects meta-analysis. *Biometrics, 55*, 732–737.

Hardy, R., Thompson, S. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine, 15*, 619–629.

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal, 41*, 901–916.

Higgins, J., Thompson, S. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Imhof, J. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika, 48*, 419–426.

Jackson, D., Bowden, J., Baker, R. (2010). How does the dersimonian and laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference, 140*, 961–970.

Knapp, G., Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*, 2693–2710.

Li, M., Liu, Y., Li, P. (2022). Empirical likelihood meta-analysis with publication bias correction under Copas-like selection model. *Annals of the Institute of Statistical Mathematics, 74*(1), 93–112.

Michael, H., Thornton, S., Xie, M. (2019). Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics, 75*, 485–493.

Noma, H., Nagashima, K., Furukawa, T. (2020). Permutation inference methods for multivariate meta-analysis. *Biometrics, 76*, 337–347.

Paule, R., Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards, 87*, 377–385.

Rover, C., Knapp, G., Friede, T. (2015). Hartung-knapp-sidik-jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC medical research methodology, 15*, 99.

Sanchez-Meca, J., Marin-Martinez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods, 13*, 31–48.

Stijnen, T., Hamza, T., Ozdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine, 29*, 3046–3067.

Thompson, S., Sharp, S. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18*, 2693–2708.

Whitehead, A., Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine, 10*, 1665–1677.