RESEARCH ARTICLE 🕕 😳



Novel approaches for random-effects meta-analysis of a small number of studies under normality

Yajie Duan¹⁰, Thomas Mathew², Demissie Alemayehu³ and Ge Cheng¹

¹Department of Statistics, Rutgers University, New Brunswick, NJ, USA

²Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD, USA

³Statistical Research and Data Science Center, Pfizer Inc., New York, NY, USA

Corresponding author: Yajie Duan; Email: yajieritaduan@gmail.com

Received: 20 August 2024; Revised: 4 May 2025; Accepted: 9 May 2025

Keywords: confidence interval; fiducial inference; modified LRT statistic; rare diseases; small sample asymptotics

Abstract

Random-effects meta-analyses with only a few studies often face challenges in accurately estimating betweenstudy heterogeneity, leading to biased effect estimates and confidence intervals with poor coverage. This issue is especially the case when dealing with rare diseases. To address this problem for normally distributed outcomes, two new approaches have been proposed to provide confidence limits of the global mean: one based on fiducial inference, and the other involving two modifications of the signed log-likelihood ratio test statistic in order to have improved performance with small numbers of studies. The performance of the proposed methods was evaluated numerically and compared with the Hartung–Knapp–Sidik–Jonkman approach and its modification to handle small numbers of studies. The simulation results indicated that the proposed methods achieved coverage probabilities closer to the nominal level and produced shorter confidence intervals compared to those based on existing methods. Two real examples are used to illustrate the proposed methods.

Highlights

What is already known?

Random-effects meta-analyses with few studies often struggle to accurately estimate between-study heterogeneity, resulting in biased effect estimates and poorly covered confidence intervals. For normally distributed outcomes, the Hartung–Knapp–Sidik–Jonkman (HKSJ) method adjusts the CI for the global mean. Rover et al. proposed a modified Knapp–Hartung (mKH) method to improve performance with a small number of studies.

What is new?

We proposed two new approaches to construct confidence intervals for the global mean: one based on fiducial inference and another using modified signed log-likelihood ratio test statistics for improved small-sample performance. Simulations showed that the new methods yielded coverage probabilities closer to nominal levels and shorter intervals than the existing methods.

Potential impact for RSM readers

Random-effects meta-analyses with a small number of studies are a common challenge, particularly in areas such as rare disease research. The proposed approaches offer readers effective alternatives with improved performance for meta-analyses with few studies, helping researchers make more informed decisions.

O O This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

[©] The Author(s), 2025. Published by Cambridge University Press on behalf of The Society for Research Synthesis Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

1. Introduction

Conducting a meta-analysis often involves a random-effects model that can account for both withinstudy and between-study variability by using a normal–normal hierarchical model.¹ For outcomes that follow a normal distribution, the data are represented as k estimates $Y_i \in \mathbb{R}$ with known within-trial variances $s_i^2 > 0$:

$$Y_i \sim N\left(\theta_i, s_i^2\right), \text{ for } i = 1, \dots, k,$$
 (1)

where the θ_i s vary from trial to trial around a global mean μ according to the normal distribution

$$\theta_i \sim N\left(\mu, \tau^2\right), \text{ for } i = 1, \dots, k.$$
(2)

Here, τ^2 is a heterogeneity variance component, representing the variability between trials or studies. The integrated and simplified model under consideration, as discussed in,¹ is

$$Y_i \sim N\left(\mu, \tau^2 + s_i^2\right), \text{ for } i = 1, \dots, k,$$
 (3)

where μ is the unknown parameter of interest. The within-trial variances s_i^2 , assumed to be known, depend on the group sizes in each trial and on the within-trial variances of the outcome variables in the *i*th trial. The latter are typically estimated, even though we proceed under the standard assumption that these are known. In many applications, only a limited number of studies is available for a meta-analysis, often fewer than five. Estimating the between-study heterogeneity τ^2 is particularly challenging in these situations. Inaccurate heterogeneity estimates can result in biased effect estimates and overly narrow confidence intervals.² Deriving reliable estimates and confidence intervals from only a small number of studies is challenging, especially in the context of rare diseases.

The problem of interest in our work is the computation of confidence limits for the underlying mean μ when the number of studies k is small. Mathes and Kuss² demonstrated that for binary outcomes and in the presence of random effects, the beta-binomial model was the most appropriate model for meta-analysis involving a small number of studies, offering a balanced trade-off between coverage probability and power. For normally distributed outcomes, there are several methods to construct confidence intervals for μ . However, there is no universally accepted satisfactory method, especially for small k. In the present work, the focus is on the outcomes following a normal distribution.

A natural point estimate of μ can be obtained as a weighted average of the Y_i s in (3), weighted by the inverse of the variance in the model (3). The resulting point estimate, say $\hat{\mu}_{\tau}$, and the associated normal distribution, are given by

$$\hat{\mu}_{\tau} = \frac{\sum_{i=1}^{k} \left(\tau^2 + s_i^2\right)^{-1} Y_i}{\sum_{i=1}^{k} \left(\tau^2 + s_i^2\right)^{-1}} \sim N\left(\mu, \hat{\sigma}_{\mu}^2 = \left\{\sum_{i=1}^{k} \left(\tau^2 + s_i^2\right)^{-1}\right\}^{-1}\right).$$
(4)

Using a normal approximation for $\hat{\mu}_{\tau}$, an approximate $(1 - \alpha)$ % confidence interval for μ is given by

$$\hat{\mu}_{\tau} \pm \hat{\sigma}_{\mu} z_{(1-\alpha/2)},\tag{5}$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ -quantile of the standard normal distribution. In practice, τ^2 is replaced with a suitable estimate, and the normal approximation is satisfactory when the number of studies k is large, with small within-trial variances or when heterogeneity is minimal.¹

An adjusted CI introduced by Hartung and Knapp,³ and Sidik and Jonkman,⁴ usually referred to as the Hartung–Knapp–Sidik–Jonkman (HKSJ) method is given by

$$\hat{\mu}_{\tau} \pm \sqrt{q} \hat{\sigma}_{\mu} t_{(k-1);(1-\alpha/2)},$$
(6)

where $q = \frac{1}{k-1} \sum_{i} \frac{(y_i - \hat{\mu}_\tau)^2}{(\tau^2 + s_i^2)}$, and $t_{(k-1);(1-\alpha/2)}$ is the $(1 - \alpha/2)$ -quantile of the Student-*t* distribution with (k - 1) degrees of freedom. For a small number of studies *k*, Rover et al.¹ proposed the modified Knapp–Hartung (mKH) method, replacing *q* by

$$q^* = \max\{1, q\}.$$
 (7)

The CI of μ using the HKSJ method in (6) is generally wider than the CI based on the normal approximation in (5) since the student-t quantile is larger than the corresponding normal quantile. However, *q* can become arbitrarily small, and if $q < \frac{z_{(1-\alpha/2)}}{t_{(k-1);(1-\alpha/2)}}$, the HKSJ method's adjusted interval would be shorter than the normal approximation-based interval, which is counter-intuitive. The mKH method provides a more conservative approach, with error rates that are more closely aligned with the normal level, particularly when dealing with a small number of studies with varying sizes or precision levels.¹

In this work, two novel approaches are proposed to construct confidence intervals for μ , and their performance is evaluated numerically through a simulation study, and compared with the existing methods. The first approach is based on the idea of fiducial inference, and the second approach is based on two modifications to the signed log-likelihood ratio test statistic to obtain accurate small sample performance.⁵ For details on fiducial inference, we refer to the works of Hannig et al.^{6,7} Earlier, fiducial quantities were investigated by Weerahandi,⁸ referring to it as generalized pivotal quantities.

The following sections first describe the proposed methods and algorithms for obtaining confidence limits of μ in (3). Next, the performance of these methods is evaluated through the results of a simulation study and two applied examples. The paper concludes with a discussion of the findings.

2. Methods

To obtain the confidence limits for μ in the model (3), especially when the number of studies k is small, two methods are proposed. One method is based on the fiducial approach, while the other involves modifications of the signed log-likelihood ratio test statistic. The fiducial approach is primarily based on the work of Iyer et al.⁹ The second method, which focuses on small sample asymptotics, includes two modifications to the signed log-likelihood ratio test to improve performance in small sample scenarios, as detailed by DiCiccio et al.⁵

2.1. The fiducial approach

Consider the point estimate of μ , denoted by $\hat{\mu}_{\tau}$, given in equation (4). Let the corresponding residual sum of squares be denoted by $R(\tau^2)$. Defining

$$\mathbf{y} = (Y_1, Y_2, \dots, Y_k)', \text{ and } \mathbf{D} = \text{diag}\left(s_1^2, s_2^2, \dots, s_k^2\right) + \tau^2 I_k,$$
 (8)

the quantity $R(\tau^2)$ can be expressed as

$$R\left(\tau^{2}\right) = (\mathbf{y} - \hat{\mu}_{\tau} \mathbf{1}_{k})' \mathbf{D}^{-1} (\mathbf{y} - \hat{\mu}_{\tau} \mathbf{1}_{k})$$

= $\mathbf{y}' \left[\mathbf{D}^{-1} - \frac{\mathbf{D}^{-1} \mathbf{1}_{k} \mathbf{1}_{k}' \mathbf{D}^{-1}}{\mathbf{1}_{k}' \mathbf{D}^{-1} \mathbf{1}_{k}} \right] \mathbf{y} \sim \chi^{2}_{k-1},$ (9)

where $\mathbf{1}_k$ is a $k \times 1$ vector of ones, and χ_r^2 denotes the central chi-square distribution with *r* degrees of freedom. The quantities $\hat{\mu}_{\tau}$ and $R(\tau^2)$ are independent, as shown by the standard results on linear models under the normality assumption.

4 Duan et al.

To understand the behavior of $R(\tau^2)$ as a function of τ^2 , consider another representation of $R(\tau^2)$. Let Z be a $k \times (k-1)$ matrix whose columns are orthonormal and orthogonal to the vector $\mathbf{1}_k$ (i.e., $\mathbf{Z'Z} = I_{k-1}$ and $\mathbf{Z'1}_k$ is a vector of zeros). Then, the following result holds:

$$\boldsymbol{D}^{-1} - \frac{\boldsymbol{D}^{-1} \mathbf{1}_k \mathbf{1}'_k \boldsymbol{D}^{-1}}{\mathbf{1}'_k \boldsymbol{D}^{-1} \mathbf{1}_k} = \mathbf{Z} \left(\mathbf{Z}' \boldsymbol{D} \mathbf{Z} \right)^{-1} \mathbf{Z}',$$

see Appendix M, Section M.4.f in the work by Searle et al.¹⁰ Thus, $R(\tau^2)$ can be represented as

$$R\left(\tau^{2}\right) = \mathbf{y}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{D}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y},$$
(10)

where **D** is the diagonal matrix given in Equation (8). From this representation, it follows that $R(\tau^2)$ is a decreasing function of τ^2 , reaching its maximum value when $\tau^2 = 0$, and is also a convex function of τ^2 . These properties were also proved by Iyer et al.⁹ based on the representation (9). Let

$$\boldsymbol{D}_0 = \operatorname{diag}\left(s_1^2, s_2^2, \dots, s_k^2\right) \tag{11}$$

which is the value of **D** in (8) when $\tau^2 = 0$. Thus, $R(\tau^2) \le R(0)$. Note that

$$R(0) = (\mathbf{y} - \hat{\mu}_0 \mathbf{1}_k)' \mathbf{D}_0^{-1} (\mathbf{y} - \hat{\mu}_0 \mathbf{1}_k)$$

= $\mathbf{y}' \left[\mathbf{D}_0^{-1} - \frac{\mathbf{D}_0^{-1} \mathbf{1}_k \mathbf{1}_k' \mathbf{D}_0^{-1}}{\mathbf{1}_k' \mathbf{D}_0^{-1} \mathbf{1}_k} \right] \mathbf{y}$
= $\mathbf{y}' \mathbf{Z} \left(\mathbf{Z}' \mathbf{D}_0 \mathbf{Z} \right)^{-1} \mathbf{Z}' \mathbf{y},$ (12)

where $\hat{\mu}_0 = \frac{\sum_{j=1}^k (s_j^2)^{-1} Y_j}{\sum_{j=1}^k (s_j^2)^{-1}}$. We note that R(0) is the usual Q- statistic to test if $\tau^2 = 0$, and $R(\tau^2)$ is the

generalized Q-statistic.¹¹

We shall now explain the derivation of a fiducial quantity for μ , whose percentiles can be used to compute confidence limits for μ . Since the fiducial quantity for μ depends on the fiducial quantity for τ^2 , we shall first obtain a fiducial quantity for the latter. Let y_o denote the observed value of y. Furthermore, let

$$R_{o}(\tau^{2}) = \mathbf{y}_{o}' \left[\mathbf{D}^{-1} - \frac{\mathbf{D}^{-1} \mathbf{1}_{k} \mathbf{1}_{k}' \mathbf{D}^{-1}}{\mathbf{1}_{k}' \mathbf{D}^{-1} \mathbf{1}_{k}} \right] \mathbf{y}_{o}$$

and
$$R_{o}(0) = \mathbf{y}_{o}' \left[\mathbf{D}_{0}^{-1} - \frac{\mathbf{D}_{0}^{-1} \mathbf{1}_{k} \mathbf{1}_{k}' \mathbf{D}_{0}^{-1}}{\mathbf{1}_{k}' \mathbf{D}_{0}^{-1} \mathbf{1}_{k}} \right] \mathbf{y}_{o}.$$
 (13)

Let $U^2 \sim \chi^2_{k-1}$. Following the work of Iyer et al.,⁹ a fiducial quantity for τ^2 is the solution to the equation $R_o(\tau^2) = U^2$, provided U^2 is between 0 and $R_o(0)$. Thus $\tilde{\tau}^2$, the fiducial quantity for τ^2 , is given by

$$\tilde{\tau}^2 = \begin{cases} \text{Solution to } R_o\left(\tau^2\right) = U^2, \text{ if } 0 < U^2 < R_o(0), \\ 0, \text{ otherwise }. \end{cases}$$
(14)

Here, we would like to note that Paule and Mandel¹² suggested an estimator of τ^2 as the solution to $R_o(\tau^2) = k - 1.$

To get a fiducial quantity for μ , note that $\hat{\mu}_{\tau}$ in (4) can be written as

$$\hat{\mu}_{\tau} = \left(\mathbf{1}_{k}^{\prime} \boldsymbol{D}^{-1} \mathbf{1}_{k}\right)^{-1} \mathbf{1}_{k}^{\prime} \boldsymbol{D}^{-1} \boldsymbol{y} \sim N\left(\mu, \left(\mathbf{1}_{k}^{\prime} \boldsymbol{D}^{-1} \mathbf{1}_{k}\right)^{-1}\right).$$

Thus, we can write

$$\mu = \hat{\mu}_{\tau} - \left(\mathbf{1}_{k}^{\prime} \boldsymbol{D}^{-1} \mathbf{1}_{k}\right)^{-\frac{1}{2}} \boldsymbol{Z}, \qquad (15)$$

where $Z = (\mathbf{1}'_k \mathbf{D}^{-1} \mathbf{1}_k)^{\frac{1}{2}} (\hat{\mu}_{\tau} - \mu) \sim N(0, 1)$. The expression on the right-hand side of the above equation depends on the unknown parameter τ^2 and the observable random variable \mathbf{y} (in the expression for $\hat{\mu}_{\tau}$). A fiducial quantity for μ is obtained from the above equation by replacing τ^2 with the fiducial quantity $\tilde{\tau}^2$ in (14) and by replacing \mathbf{y} with its observed value \mathbf{y}_o . Thus, let

$$\tilde{\boldsymbol{D}} = \operatorname{diag}\left(s_1^2, s_2^2, \dots, s_k^2\right) + \tilde{\tau}^2 I_k, \quad \text{and} \quad \hat{\mu}_{\tilde{\tau}} = \left(\mathbf{1}'_k \tilde{\boldsymbol{D}}^{-1} \mathbf{1}_k\right)^{-1} \mathbf{1}'_k \tilde{\boldsymbol{D}}^{-1} \mathbf{y}_o.$$
(16)

A fiducial quantity for μ , denoted by $\tilde{\mu}$, is then given by

$$\tilde{\mu} = \hat{\mu}_{\tilde{\tau}} - \left(\mathbf{1}_k' \tilde{\boldsymbol{D}}^{-1} \mathbf{1}_k\right)^{-\frac{1}{2}} Z.$$
(17)

We note that $\tilde{\mu}$ is a function of two independent random variables: the chi-square random variable U^2 with df = k - 1 used in the computation of $\tilde{\tau}^2$ in (14), and a standard normal random variable Z. Once the observed data vector \mathbf{y}_o is available, a large number of values, say *m*, of the pair (U^2, Z) can be generated. These *m* pairs of values, along with the observed data \mathbf{y}_o , can be used to compute *m* realizations of the fiducial quantities $\tilde{\tau}^2$ and $\tilde{\mu}$. The percentiles of the *m* realizations of $\tilde{\mu}$ provide confidence limits for μ .

In the above approach, point estimates of μ and τ^2 are not explicitly used. An algorithm for implementing the fiducial approach to obtain confidence limits for μ is outlined below.

Algorithm 1 Fiducial Approach

Require: Observed data \mathbf{y}_o , and known within-trial variances s_i^2 , $i = 1, \dots, k$. Sample *m* pairs of (U^2, Z) from $U^2 \sim \chi_{k-1}^2$, and $Z \sim N(0, 1)$. **for** each pair of (U^2, Z) **do if** $0 < U^2 < R_o(0)$ **then** Compute $\tilde{\tau}^2$ by solving $R_o(\tau^2) = U^2$. **else** Set $\tilde{\tau}^2 = 0$ **end if** Compute $\tilde{\mu} = -\hat{\mu}_{\tilde{\tau}} - (\mathbf{1}'_k \tilde{\boldsymbol{D}}^{-1} \mathbf{1}_k)^{-\frac{1}{2}} Z$, with $\tilde{\boldsymbol{D}} = -\text{diag}(s_1^2, s_2^2, \dots, s_k^2) + \tilde{\tau}^2 I_k$, and $\hat{\mu}_{\tilde{\tau}} = (\mathbf{1}'_k \tilde{\boldsymbol{D}}^{-1} \mathbf{1}_k)^{-1} \mathbf{1}'_k \tilde{\boldsymbol{D}}^{-1} \mathbf{y}_o$. **end for return** *m* realizations of the fiducial quantity $\tilde{\mu}$, with their percentiles as confidence limits for μ .

2.2. Small sample asymptotics

It is well known that under regularity conditions, the signed log-likelihood ratio test statistic is asymptotically standard normal, providing first-order accuracy. Small-sample asymptotics involve

6 Duan et al.

modifying the signed log-likelihood ratio test statistic to achieve accurate performance in small samples. In order to introduce the basic idea, consider an $n \times 1$ vector of observations y whose distribution depends on a $d \times 1$ parameter vector θ . Let $\ell(\theta) = \ell(\theta; y)$ be the log-likelihood function. Suppose that the parameter θ is partitioned as $(\psi, \lambda')'$, where ψ is a scalar parameter of interest, and λ is a nuisance parameter. Let $\hat{\theta} = (\hat{\psi}, \hat{\lambda}')'$ denote the MLE, $\hat{\lambda}_{\psi}$ denote the constrained MLE of λ for a fixed ψ , and write $\hat{\theta}_{\psi} = (\psi, \hat{\lambda}'_{\psi})'$. Large sample inference for ψ can be based on the signed log-likelihood root

$$r(\psi) = \operatorname{sign}(\hat{\psi} - \psi) \left[2 \left\{ \ell(\hat{\theta}) - \ell(\hat{\theta}_{\psi}) \right\} \right]^{1/2},$$
(18)

where sign(x) is +1 if x > 0 and -1 if x < 0. Under regular conditions, $r(\psi)$ is asymptotically standard normal with an error of the order $O(n^{-1/2})$. Several modifications of $r(\psi)$ are available in the literature to improve small sample performance.¹³ Here, we consider two modifications proposed by DiCiccio et al.⁵

2.2.1. Modification I

To describe the first modification, let $E(r(\psi)) = m(\psi, \lambda)$ and $Var(r(\psi)) = 1 + v(\psi, \lambda)$. Now define

$$r_1(\psi) = \frac{r(\psi) - m\left(\psi, \hat{\lambda}_{\psi}\right)}{\left\{1 + v\left(\psi, \hat{\lambda}_{\psi}\right)\right\}^{1/2}}.$$
(19)

For a fixed value of ψ , $r_1(\psi)$ is thus a standardization of $r(\psi)$ using the mean and variance evaluated at $\lambda = \hat{\lambda}_{\psi}$. DiCiccio et al.⁵ have shown that under regular conditions, $r_1(\psi)$ has an asymptotic standard normal distribution, and $P[r_1(\psi) \ge t] = [1 - \Phi(t)] + O(n^{-3/2})$. In other words, the tail area approximation based on the asymptotic standard normal distribution of $r_1(\psi)$ is significantly more accurate than that based on the asymptotic standard normal distribution of $r(\psi)$.

In order to carry out inference based on the asymptotic standard normal distribution of $r_1(\psi)$, it is necessary to compute the mean and variance of $r(\psi)$; namely, the quantities $m\left(\psi, \hat{\lambda}_{\psi}\right)$ and $1 + v\left(\psi, \hat{\lambda}_{\psi}\right)$. For a specified value of ψ , these can be obtained numerically, as noted in the work of DiCiccio et al.⁵ After specifying ψ , compute $\hat{\lambda}_{\psi}$, and generate a parametric bootstrap sample of size *n* when ψ takes the specified value and $\lambda = \hat{\lambda}_{\psi}$. Let $r^*(\psi)$ denote the value of $r(\psi)$ calculated from the parametric bootstrap sample. Generate several parametric bootstrap samples of size *n*, say *B* samples, and compute *B* values of $r(\psi)$, denoted by $r_j^*(\psi)$ for j = 1, 2, ..., B. The mean and variance of $r_j^*(\psi)$ for j = 1, 2, ..., B, provide estimates of $m\left(\psi, \hat{\lambda}_{\psi}\right)$ and $1 + v\left(\psi, \hat{\lambda}_{\psi}\right)$, respectively, to be used in the calculation of $r_1(\psi)$. To compute confidence limits for ψ based on $r_1(\psi)$, equate $r_1(\psi)$ to standard normal percentiles and solve for ψ . For example, a 95% upper confidence limit for ψ can be obtained as the solution for ψ to equation $r_1(\psi) = -1.96$.

2.2.2. Modification II

A second modification suggested by DiCiccio et al.⁵ involves estimating the tail probability $P(r(\psi) \le t)$ using Monte Carlo simulation at the parameter value $\theta = \hat{\theta}_{\psi}$, rather than relying on the standard normal distribution for $r(\psi)$. For example, to test the null hypothesis $\psi = \psi_0$ against the one-sided alternative $\psi > \psi_0$, a *p*-value can be calculated as $P(r(\psi_0) \le r_{obs})$, evaluated at the parameter value $\theta_0 = \hat{\theta}_{\psi_0}$, where r_{obs} is the observed value of $r(\psi_0)$. The confidence limits for ψ can then be obtained by inverting the test.

2.2.3. Algorithms

In the context of the random-effects meta-analysis model given in (3), the parameter of interest is $\psi = \mu$, and the nuisance parameter is the scalar $\lambda = \tau^2$. Thus, in terms of the notation used for the two modifications given above, the parameter vector is $\theta = (\mu, \tau^2)'$. We use $\ell(\theta) = \ell(\mu, \tau^2)$ to denote the log-likelihood function under the model (3).

The algorithms proposed to implement the two modifications to obtain the confidence limits of μ under the model (3) are presented below.

Algorithm 2 Small Sample Asymptotics with Modification I

Require: Observed data y_o , and known within-trial variances s_i^2 , $i = 1, \dots, k$.

- 1: Consider a grid of values of μ_0 .
- 2: for each value μ_0 do
- 3: Compute the MLEs $\hat{\mu}$ and $\hat{\tau}^2$ by maximizing the likelihood function $\ell(\mu, \tau^2)$.
- 4: Let $\mu = \mu_0$, maximize the likelihood function $\ell(\mu_0, \tau^2)$ with respect to τ^2 , i.e., obtain $\hat{\tau}^2_{\mu_0}$.
- 5: Compute $r(\mu_0) = \text{sign}(\hat{\mu} \mu_0) \left[2 \left\{ \ell(\hat{\mu}, \hat{\tau}^2) \ell(\mu_0, \hat{\tau}_{\mu_0}^2) \right\} \right]^{1/2}$
- 6: **for** each $j = 1, 2, \dots, B$ **do**
- 7: Generate parametric bootstrap sample from $Y_{ij}^* \sim N\left(\mu_0, \hat{\tau}_{\mu_0}^2 + s_i^2\right), i = 1, \dots, k$
- 8: Repeat steps 3-5 for Y_{ii}^* , $i = 1, \dots, k$, and obtain $r_i^*(\mu_0)$.
- 9: end for

10: Compute the mean and variance of $r_j^*(\mu_0)$, $j = 1, 2, \cdots, B$, denoted by $m\left(\mu_0, \hat{\tau}_{\mu_0}^2\right)$ and $1 + v\left(\mu_0, \hat{\tau}_{\mu_0}^2\right)$,

then compute
$$r_1(\mu_0) = \frac{r(\mu_0) - m(\mu_0, \hat{\tau}^2_{\mu_0})}{1 + v(\mu_0, \hat{\tau}^2_{\mu_0})^{1/2}}$$

12: **return** Confidence limits for μ are the values of μ_0 for which $r_1(\mu_0)$ is equal to the appropriate standard normal quantiles.

Algorithm 3 Small Sample Asymptotics with Modification II

Require: Observed data y_o , and known within-trial variances s_i^2 , $i = 1, \dots, k$.

- 1: Consider a grid of values of μ_0 .
- 2: for each value μ_0 do
- 3: Compute the MLEs $\hat{\mu}$ and $\hat{\tau}^2$ by maximizing the likelihood function $\ell(\mu, \tau^2)$.
- 4: Let $\mu = \mu_0$, maximize the likelihood function $\ell(\mu_0, \tau^2)$ with respect to τ^2 , i.e., obtain $\hat{\tau}^2_{\mu_0}$

5: Compute
$$r(\mu_0) = \text{sign}(\hat{\mu} - \mu_0) \left[2 \left\{ \ell(\hat{\mu}, \hat{\tau}^2) - \ell(\mu_0, \hat{\tau}_{\mu_0}^2) \right\} \right]^1$$

- 6: **for** each $j = 1, 2, \dots, B$ **do**
- 7: Generate parametric bootstrap sample from $Y_{ij}^* \sim N\left(\mu_0, \hat{\tau}_{\mu_0}^2 + s_i^2\right), i = 1, \dots, k$
- 8: Repeat steps 3-5 for Y_{ij}^* , $i = 1, \dots, k$, and obtain r_j^* (μ_0).
- 9: end for
- 10: Compute $P(r(\mu_0)^* \le r(\mu_0))$ and $P(r(\mu_0)^* \ge r(\mu_0))$ as the corresponding frequencies in the bootstrap samples.
- 11: end for
- 12: return Confidence limits for μ are the values of μ_0 for which the p-values are equal to the significance level.

For the numerical implementation of the proposed methods, a grid of values for μ_0 can be considered for both algorithms. The grid of values μ can be selected to cover a range around the MLE of μ , since the upper confidence limit will be larger than the MLE of μ , and the lower confidence limit will be smaller. For the algorithm for modification I, to calculate a two-sided confidence interval for μ , the upper and lower confidence limits are obtained by finding the values of μ_0 for which $r_1(\mu_0)$ is close to the standard normal quantiles according to a specified threshold. For the algorithm for modification II, the confidence limits for μ are determined by finding the values of μ_0 that make the *p* values $P(r(\mu_0)^* \le r(\mu_0))$ and $P(r(\mu_0)^* \ge r(\mu_0))$ close to the significance level, once again according to a specified threshold.

3. Results

3.1. Simulation settings

In order to evaluate the performance of the proposed methods, a simulation study similar to that of Rover et al.¹ was conducted for the modified HKSJ (mKH) method with a small number of studies (k). The "relative amount of heterogeneity" is expressed in terms of the measure I^2 defined as

$$I^{2} = \frac{\tau^{2}}{\tau^{2} + \tilde{s}^{2}},$$
(20)

where \tilde{s} is an "average" standard error among the study-specific s_i values.^{1,14} In the simulation study, \tilde{s}^2 was calculated as the arithmetic mean of the squared standard errors, i.e.,

$$\tilde{s}^2 = \frac{1}{k} \sum_{i=1}^k \frac{2\sigma^2}{n_i}$$
(21)

with n_i being the group size of trial i (i = 1, ..., k) and σ is a scaling factor. Since the results depend solely on the ratio τ/σ , $\sigma = 1$ was set in the simulations.¹⁵ The simulation setup we have used and also adopted in the work of Rover et al.,¹ is taken from the study by IntHout et al.¹⁵

Similarly to the work by Rover et al.,¹ four different meta-analyses scenarios were considered in the simulation: (A) trials of equal size, (B) equally sized trials but including one small trial, (C) with half large and half small trials, and (D) equally sized trials and one large trial. A large trial is 10 times the size of a small trial, which means that the associated standard errors differ by approximately a factor of 3.¹ The number of trials *k* ranged from 2 to 10, that is, k = 2, ..., 10. The true level of heterogeneity was set at $I^2 = 0.5$; see (20). The average group size n_i in each trial was set at around 100. Without loss of generality, the true value of μ was set to zero. The 95% confidence interval (CI) of μ was constructed using the proposed methods and compared to the HKSJ method³ and the modified HKSJ (mKH) method for small k.¹

The simulation for normally distributed outcomes under the model (3) involved several key steps, as detailed in Appendix 2 of the work of IntHout et al.¹⁵ For each scenario with $I^2 = 0.5$, the variance τ^2 was calculated using equations (20) and (21). For each trial *i*, the 'true' trial-specific effect size θ_i in (2) was drawn from a normal distribution with mean $\mu = 0$ and variance τ^2 . The trial outcomes were then generated from a normal distribution with mean θ_i and variance $\frac{2\sigma^2}{n_i} = 2/n_i$ (recall that the scaling factor σ is chosen to be equal to one). Additionally, for each trial *i* with group size n_i , the variance of the trial outcome, i.e., the quantity s_i^2 , was generated based on a χ^2 distribution with $2n_i - 2$ degrees of freedom, divided by $(n_i - 1)n_i$; see the work of IntHout et al.¹⁵ For each simulation, the generated observed trial outcomes y_o and within-trial variance s_i^2 for $i = 1, \ldots, k$, were used as input for the proposed algorithms to construct 95% CIs of μ . For the fiducial approach, 5,000 realizations of the fiducial quantities were used to obtain the confidence limits. For asymptotic modifications I & II, 1000 bootstrap samples were used, i.e., B = 1,000.

To compare the proposed methods with the existing ones, the CIs of μ were also constructed using the HKSJ method³ and the modified HKSJ (mKH) method.¹ The DerSimonian–Laird (DL) estimator¹⁶ of the between-study variance τ^2 was used for both the HKSJ and mKH methods in the simulation. According to Rover et al.¹ and Sánchez-Meca and Marín-Martínez,¹⁷ the constructed CIs for both the



Figure 1. Plots of the 95% CI coverage, median and 90th quantiles of the CI width ratios compared to the HKSJ method for trials with half large/small sizes and average size around 100 (with a ceiling (k/2) of small sizes for kodd), by Asymptotic Modification I (Asym_I) method, Asymptotic Modification II (Asym_II) method, Fiducial approach, HKSJ method, and modified HKSJ (mKH) method.

HKSJ and mKH methods are similar if using the DL estimator, the restricted maximum likelihood (REML) estimator,¹⁸ and the Paule–Mandel (PM) estimator.¹²

A total of 5,000 meta-analyses were simulated, and all simulations were performed using R.¹⁹ In order to compare the different approaches for computing the confidence intervals for μ , the coverage probabilities and the expected widths of the CIs were calculated, where the latter was expressed as a ratio to the expected length of the interval obtained by the HKSJ method. The median of these ratios and the 90th percentile of the ratios were also evaluated, as in the work of Rover et al.¹

3.2. Simulation results

For trials with half large and half small sizes, with the average size around 100, the simulation results are shown in Figure 1. The results include 95% CI coverage, median and 90th quantiles of the expected CI length ratios mentioned above, for the number of trials k = 2, ..., 10. A large trial was 10 times the size of a small trial, with the overall trial average being around 100, and the number of small-sized trials was chosen to be (k + 1)/2 or k/2 depending on whether k is odd or even. For example, for k = 3, the group sizes for the trials would be n = (25, 25, 250); for k = 2, the group sizes would be n = (18, 180).



Figure 2. Plots of the 95% CI coverage, median and 90th quantiles of the CI width ratios compared to the HKSJ method for trials with equal sizes ofn = 100, by Asymptotic Modification I (Asym_I) method, Asymptotic Modification II (Asym_II) method, Fiducial approach, HKSJ method, and modified HKSJ (mKH) method.

The simulation results demonstrate that the proposed new methods, including the fiducial approach and the small sample asymptotic methods, particularly Asymptotic Modification II, perform better than the modified HKSJ (mKH) method when the number of trials is small. The HKSJ method underperforms for small k, and while its performance is indeed enhanced by the modified HKSJ (mKH) method, the proposed new methods offer superior results. The fiducial approach consistently provides more accurate confidence interval (CI) coverage. Asymptotic Modification II, in particular, shows more accurate CI coverage and shorter CI lengths for small numbers of trials (k = 2, 3), compared to other methods. Additionally, Asymptotic Modification I also offers a more accurate CI coverage when k > 4. The fiducial approach yields larger median CI lengths, but with less variability in CI lengths, especially for k = 2, 3. On the other hand, Asymptotic Modification I results in CI lengths similar to mKH but with a larger median, and Asymptotic Modification II shows smaller CI lengths for k = 2 and less variability of CI lengths for k = 3, 4.

Figure 2 displays the simulation results for trials with equal group sizes of 100. In this scenario, the HKSJ method can handle small k effectively. It works very well when the analyzed studies are of equal size (i.e., have equal standard errors), as shown in the work of Rover et al.¹

Figures 3 and 4 show the corresponding simulation results for the scenarios with one large trial, and with one small trial, respectively, with an average group size around 100. The results and conclusions



Figure 3. Plots of the 95% CI coverage, median and 90th quantiles of the CI width ratios compared to the HKSJ method for the case of equally-sized trials with one large trial and average size around 100, by Asymptotic Modification I (Asym_I) method, Asymptotic Modification II (Asym_II) method, Fiducial approach, HKSJ method, and modified HKSJ (mKH) method.

are similar as in the case of trials with half-large and half-small sizes. The HKSJ method underperforms for small k, while the mKH method improves its performance. The fiducial approach consistently provides more accurate CI coverage close to the nominal 95%. Asymptotic Modification II maintains more accurate CI coverage, outperforming the mKH method for small numbers of trials. For the CI lengths, the results reveal that the fiducial approach yields larger median CI lengths with less variability, while Asymptotic Modification II shows smaller and more stable CI lengths for small k, indicating its robustness.

In summary, for computing a CI of μ in unbalanced settings with a small number of studies k, Asymptotic Modification I shows results comparable to those based on existing methods. However, the fiducial approach and Asymptotic Modification II show better results, with more accurate CI coverages and shorter CI lengths. The findings highlight the superior performance of the newly proposed methods in terms of both CI coverage and length, making them excellent choices for practical applications.

4. Examples

An alternative methodology, not focused on in our work, is a Bayesian approach (see the works of Bender et al.,²⁰ and Friede et al.,²¹) which introduces stability through the incorporation of prior



Figure 4. Plots of the 95% CI coverage, median and 90th quantiles of the CI width ratios compared to the HKSJ method for the case of equally-sized trials with one small trial and average size around 100, by Asymptotic Modification I (Asym_I) method, Asymptotic Modification II (Asym_II) method, Fiducial approach, HKSJ method, and modified HKSJ (mKH) method.

knowledge about the heterogeneity parameter τ^2 through (weakly) informative prior distributions. The two examples of few studies illustrated in the work of Bender et al.²⁰ are used here to demonstrate the performance of the proposed methods. These examples come from dossier assessments conducted by the Institute for Quality and Efficiency in Health Care.²⁰ Assuming that the logarithm of hazard ratios (HR) or relative risks (RR) follows a normal distribution, 95% confidence intervals for the data in both examples were generated using the HKSJ method, the modified HKSJ (mKH) method, the Bayesian approach, and the proposed methods, including the Fiducial approach and Asymptotic Modification I and II. For Bayesian approaches, half-normal priors for the heterogeneity parameter τ with scales of 0.5 (Bayesian-HN(0.5)) and 1.0 (Bayesian-HN(1)) were applied.

In the first example, the added benefit of belatacept compared to ciclosporin A in combination with corticosteroids and mycophenolate mofetil was assessed as the appropriate comparator therapy for the prophylaxis of graft rejection in adults receiving a renal transplant.²⁰ Only k = 2 studies were available. The results for this example are illustrated in Figure 5. The HKSJ method yielded a wide 95% CI due to insufficient data to reliably estimate heterogeneity, leading to a conservative result. The mKH method produced the same result as HKSJ in this case. The Bayesian approach provides an alternative, yielding narrower intervals than HKSJ and mKH. However, prespecification of the prior distribution for between-study variation τ^2 is crucial to interpreting the intervals in a frequentist

BENEFIT BENEFIT-EXT		⊢⊷⊣ ⊦∎₁		0.44 [0.32, 0.61] 0.60 [0.47, 0.77]
HKSJ mKH Bayesian-HN(0.5) Bayesian-HN(1) Fiducial Asym_I Asym_I	0.05		5	0.52 [0.07, 3.70] 0.52 [0.07, 3.70] 0.53 [0.27, 0.98] 0.53 [0.17, 1.52] 0.51 [0.10, 2.72] 0.52 [0.18, 1.48] 0.52 [0.17, 1.48]

Figure 5. A forest plot showing random-effects meta-analysis results for the first belatacept example, by HKSJ method, modified HKSJ (mKH) method, Bayesian method using half-normal priors fortwith scales 0.5 (Bayesian-HN(0.5)) and 1 (Bayesian-HN(1)), Fiducial approach, and Asymptotic Modification I (Asym_I) and II (Asym_II) methods.

IMPACT D9901 D9902A		- ₽ -1 		2.14 [1.41, 3.24] 7.68 [1.92, 30.76] 3.02 [0.97, 9.44]
HKSJ mKH Bayesian-HN(0.5) Bayesian-HN(1) Fiducial Asym_I Asym_I				2.87 [0.70, 11.84] 2.87 [0.70, 11.84] 2.60 [1.40, 6.34] 2.64 [1.13, 9.50] 3.01 [0.71, 16.89] 2.87 [0.68, 15.20] 2.87 [0.80, 14.46]
	0.1	1 RR	10	100

Figure 6. A forest plot showing random-effects meta-analysis results for the second sipuleucel-T example, by HKSJ method, modified HKSJ (mKH) method, Bayesian method using half-normal priors fortwith scales 0.5 (Bayesian-HN(0.5)) and 1 (Bayesian-HN(1)), Fiducial approach, and Asymptotic Modification I (Asym_I) and II (Asym_II) methods.

sense, as a hypothesis test.²⁰ The application of a half-normal prior with scale 0.5 for τ results in a statistically significant pooled effect estimate, while a scale of 1 does not. Among the proposed methods, Asymptotic Modifications I and II generate similar but slightly shorter confidence intervals compared to the Bayesian method with a half-normal prior of scale 1. The Fiducial approach provides slightly wider intervals than the asymptotic modification methods but remains narrower than those from HKSJ and mKH.

In the second example, the added benefit of sipuleucel-T compared to the appropriate comparator therapy to treat asymptomatic or minimally symptomatic metastatic (non-visceral) castrate-resistant prostate cancer was assessed in male adults for whom chemotherapy is not yet clinically indicated.²⁰ Here, k = 3 relevant studies were available. The results for this example are shown in Figure 6. Bayesian approaches yield shorter confidence intervals, indicating a statistically significant difference between treatment groups, to the disadvantage of sipuleucel-T. The mKH method produces the same wide confidence interval as HKSJ, including the relative risk of 1. The three proposed methods generate wider confidence intervals, particularly on the right-hand side, covering the finding from the second

study with a large upper confidence limit for the relative risk, which may suggest better coverage of the true value. Although Bayesian approaches tended to give qualitatively different results relative to the other methods, the choice of method generally had negligible impact on the conclusions drawn in this example.

5. Discussion

In this work, we have proposed two novel methods for computing confidence limits for μ in randomeffects meta-analysis, particularly suited for a small number of studies. One method is based on fiducial inference, while the other involves two small sample asymptotic procedures, referred to as Asymptotic Modifications I and II. Our simulation studies indicate that for unbalanced settings, when the number of studies k is small, the proposed methods show significant improvements compared to existing methods. The fiducial approach and Asymptotic Modification II exhibit superior performance, providing more accurate CI coverages and shorter CI lengths compared to those based on the modified HKSJ method proposed by Rover et al.¹ Also, Asymptotic Modification I provides results that are comparable to those based on the modified HKSJ method. We believe that these findings have important practical implications since meta-analyses with a small number of studies are frequently encountered in applications. It is noted that for k = 2, none of the competing approaches yields useful information, so the differences in relative performance of the approaches do not have practical relevance.

For methods based on small sample asymptotics, other modifications of the statistic $r(\psi)$ given in (18) are also available in the literature, with the goal of achieving improved small sample performance; see the work of Brazzale et al.¹³ Such modifications have in fact been applied in the context of the model and problem investigated in our work; see the work of Guolo.²² Figures 1-3 in the work of Guolo²² show that such modifications exhibit a performance quite similar to what we have noted for Asymptotic Modifications I and II considered in our investigation. We also want to point out that in the context of inter-laboratory studies where the model (3) is assumed, but with unknown within-trial variances, the fiducial approach was developed in the work of Tian²³ and small sample asymptotic procedures were investigated by Sharma and Mathew.²⁴ Furthermore, Zejnullahi and Hedges²⁵ for small meta-analyses, evaluating degrees-of-freedom adjustments for confidence intervals.

In addition to the methodological contributions, we are developing an R package that will offer a practical tool for the easy implementation of the proposed methods.

Future research could explore the impact of estimated within-trial variability on the performance of the CIs of μ . Furthermore, the median or mean of the fiducial quantities could be used as a point estimate of μ . The performance of such a point estimate of μ could be evaluated and compared with the maximum likelihood estimate (MLE), the restricted MLE (REML), and the estimate resulting from the DerSimonian–Laird (DL) approach. These are currently under investigation.

Acknowledgements. Our sincere thanks to Professor Javier Cabrera of the Department of Statistics at Rutgers University for his valuable discussions and suggestions on statistical methods and simulation study design.

Author contributions. Conceptualization: T.M., D.A.; Methodology and Investigation: T.M., D.A., Y.D., G.C.; Formal analysis, Data curation, Software and Visualization: Y.D., G.C.; Supervision and Validation: T.M., D.A.; Writing—original draft: Y.D., T.M., G.C.; Writing—review and editing: T.M., D.A., Y.D., G.C.

Competing interest statement. The authors declare that no competing interests exist.

Data availability statement. Program codes, simulated data summary, and real data examples supporting the findings of this study can be found in the Supplementary Material.

Funding statement. The authors declare that no specific funding has been received for this article.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/rsm.2025.10022.

References

- Rover C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects metaanalysis with few studies. *BMC Med. Res. Methodol.* 2015;15: 99.
- [2] Mathes T, Kuss O. A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Res. Synth. Methods* 2018;9(3): 366–381.
- [3] Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat. Med.* 2001;20(24): 3875–3889.
- [4] Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. Stat. Med. 2002;21(21): 3253–3315.
- [5] DiCiccio TJ, Martin MA, Stern SE. Simple and accurate one-sided inference from signed roots of likelihood ratios. Can. J. Stat. 2001;29: 69–76.
- [6] Hannig J. On generalized fiducial inference. Stat. Sinica 2009;19: 491-544.
- [7] Hannig J, Iyer H, Lai RC, Lee TC. Generalized fiducial inference: A review and new results. J. Amer. Stat. Assoc. 2016;111: 1346–1361.
- [8] Weerahandi S. Generalized confidence intervals. J. Amer. Stat. Assoc. 1993;88: 899-905.
- [9] Iyer HK, Wang CMJ, Mathew T. Models and confidence intervals for true values in interlaboratory trials. J. Amer. Stat. Assoc. 2004;99: 1060–1071.
- [10] Searle SR, Casella G, McCulloch CE. Variance Components. John Wiley; 1992.
- [11] Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. Stat. Med. 2007;26: 37-52.
- [12] Paule RC, Mandel J. Consensus values and weighting factors. J. Res. Nat. Bureau Stand. 1982;87(5): 377.
- [13] Brazzale AR, Davison AC, Reid N. Applied Asymptotics: Case Studies in Small-Sample Statistics. Cambridge University Press; 2007.
- [14] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat. Med. 2002;21(11): 1539–1558.
- [15] IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med. Res. Methodol. 2014;14: 1–12.
- [16] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control. Clin. Trials 1986;7(3): 177-188.
- [17] Sánchez-Meca J, Marín-Martínez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychol. Methods* 2008;13(1): 31.
- [18] Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. J. Educ. Behav. Stat. 2005;30(3): 261–293.
- [19] R Core Team R, et al. R: A language and environment for statistical computing. R foundation for statistical computing. 2013.
- [20] Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. *Res. Synth. Methods* 2018;9: 382–392.
- [21] Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biom. J.* 2017;59(4): 658–671.
- [22] Guolo A. Higher-order likelihood inference in meta-analysis and meta-regression. Stat. Med. 2012;31: 313–327.
- [23] Tian L. Generalized inferences on the overall treatment effect in meta-analysis with normally distributed outcomes. *Biom. J.* 2008;50: 237–247.
- [24] Sharma G, Mathew T. Higher order inference for the consensus mean in inter-laboratory studies. *Biom. J.* 2011;53: 128–136.
- [25] Zejnullahi R, Hedges LV. Robust variance estimation in small meta-analysis with the standardized mean difference. *Res. Synth. Methods* 2024;15(1): 44–60.

Cite this article: Duan Y, Mathew T, Alemayehu D, Cheng G. Novel approaches for random-effects meta-analysis of a small number of studies under normality. *Research Synthesis Methods*. 2025;0: 1–15. https://doi.org/10.1017/rsm.2025.10022