On random-effects meta-analysis

D. ZENG and **D. Y. LIN**

Department of Biostatistics, CB #7420, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A

D. ZENG: dzeng@bios.unc.edu; D. Y. LIN: lin@bios.unc.edu

Summary

Meta-analysis is widely used to compare and combine the results of multiple independent studies. To account for between-study heterogeneity, investigators often employ random-effects models, under which the effect sizes of interest are assumed to follow a normal distribution. It is common to estimate the mean effect size by a weighted linear combination of study-specific estimators, with the weight for each study being inversely proportional to the sum of the variance of the effect-size estimator and the estimated variance component of the random-effects distribution. Because the estimator of the variance component involved in the weights is random and correlated with study-specific effect-size estimators, the commonly adopted asymptotic normal approximation to the meta-analysis estimator is grossly inaccurate unless the number of studies is large. When individual participant data are available, one can also estimate the mean effect size by maximizing the joint likelihood. We establish the asymptotic properties of the meta-analysis estimator and the joint maximum likelihood estimator when the number of studies is either fixed or increases at a slower rate than the study sizes and we discover a surprising result: the former estimator is always at least as efficient as the latter. We also develop a novel resampling technique that improves the accuracy of statistical inference. We demonstrate the benefits of the proposed inference procedures using simulated and empirical data.

Keywords

Clustered data; Evidence-based medicine; Genetic association; Heterogeneity; Individual patient data; Maximum likelihood estimation; Random-effects model; Research synthesis; Summary statistic

1. Introduction

Meta-analysis compares and combines results from multiple independent studies, with the hope of identifying consistent patterns and sources of disagreement. It has been routinely used in all areas of statistical applications, from astronomy to zoology. The meta-analysis literature has grown exponentially over the past three decades, owing to the need for reliable summarization of the vast and expanding volume of scientific research. According to the Web of Science, there were 2006 meta-analysis publications during the years 1985–1994, 13

154 during 1995–2004, and more than 55 000 during 2005–2014. The impact of meta-analysis on medical research and clinical practice has been enormous (see, e.g., Whitehead, 2002; Sutton & Higgins, 2008). Most of the recent discoveries about genetic variants associated with complex human diseases and traits have been made through meta-analysis (e.g., McCarthy et al., 2008; Evangelou & Ioannidis, 2013). Meta-analysis is likely to play an even greater role in the current big-data era.

Conventional meta-analysis relies on summary statistics, such as estimated effect sizes and standard errors, from relevant studies. Advances in technology and communications have made it increasingly feasible to access data on individual participants (e.g., Sutton et al., 2000). Indeed, joint analysis of individual patient data is considered the gold standard in systematic reviews of randomized clinical trials (e.g., Chalmers et al., 1993). Recently, a number of consortia have been formed to share individual participant data from genetic association studies (e.g., Psychiatric GWAS Consortium Steering Committee, 2009; Evangelou & Ioannidis, 2013; Lin et al., 2013).

Suppose that there are K independent studies, with n_k participants for the kth study. The data consist of $\{\mathcal{O}_{ki} : k = 1,..., K; i = 1,..., n_k\}$, where $\mathcal{O}_{ki}$ represents the observation, including the response variable and explanatory variables, on the ith participant of the kth study. Assume that the density function of $\mathcal{O}_{ki}$ is proportional to $f_k(\cdot; \beta_k, \eta_k)$, where β_k is the parameter of interest, namely the effect size for a new treatment or genetic mutation, and η_k is a set of nuisance parameters, such as the error variance and the regression effects of demographic variables. Let $\hat{\beta}_k$ be the maximum likelihood estimator of β_k based on the likelihood $L_k(\beta_k, \eta_k) = \prod_{i=1}^{n_k} f_k(\mathcal{O}_{ki}; \beta_k, \eta_k)$, and let $\hat{V}_k$ be the estimated variance of $\hat{\beta}_k$. We assume that standard regularity conditions (Cox & Hinkley, 1974, p. 281) hold and that the n_k are large enough that $\hat{\beta}_k$ is approximately normal with mean β_k and variance $\hat{V}_k$.

Under the fixed-effects model, $\beta_k = \beta$ for all $k = 1,..., K$. The familiar inverse-variance estimator of β is

$$\hat{\beta} = \left(\sum_{k=1}^{K} \hat{V}_k^{-1} \right)^{-1} \sum_{k=1}^{K} \hat{V}_k^{-1} \hat{\beta}_k,$$

which is approximately normal with mean β and variance $\left(\sum \hat{V}_k^{-1} \right)^{-1}$. Let $\hat{\beta}^\dagger$ be the maximum likelihood estimator of β found by maximizing the joint likelihood $\prod_{k=1}^{K} L_k(\beta, \eta_k)$. Olkin & Sampson (1998) and Mathew & Nordstrom (1999) showed that, in the case of comparing multiple treatments and a control for a continuous outcome with known error variances, $\hat{\beta}$ is the same as $\hat{\beta}^\dagger$. Recently, Lin & Zeng (2010) proved that the two estimators are asymptotically equivalent for all commonly used parametric and semiparametric models.

We focus on the random-effects model

$$\beta_k = \beta + \xi_k \quad (k=1,\ldots,K),$$

where $\xi_k \sim N(0,\ \tau^2)$. Because $\hat{\beta_k}$ is approximately normal with mean β and variance $V_k + \tau^2$, it is natural to estimate β by

$$\hat{\beta}_{\mathrm{MA}} = \left(\sum_{k=1}^{K} \hat{w}_k\right)^{-1} \sum_{k=1}^{K} \hat{w}_k \hat{\beta}_k, \quad (1)$$

where $\hat{w}_k = (\hat{V}_k + \hat{\tau}^2_{\mathrm{MA}})^{-1}$, with (DerSimonian & Laird, 1986)

$$\hat{\tau}^2_{\mathrm{MA}} = \max\left\{0, \frac{\sum_{k=1}^{K} \hat{V}_k^{-1}(\hat{\beta}_k - \hat{\beta})^2 - (K-1)}{\sum_{k=1}^{K} \hat{V}_k^{-1} - \sum_{k=1}^{K} \hat{V}_k^{-2}/\sum_{k=1}^{K} \hat{V}_k^{-1}}\right\}. \quad (2)$$

Inference on β is typically based on the normal approximation

$$\hat{\beta}_{\mathrm{MA}} \sim N\left\{\beta,\ \left(\sum \hat{w}_k\right)^{-1}\right\}. \quad (3)$$

For finite K, the estimator $\hat{\tau}^2_{\mathrm{MA}}$ involved in the $\hat{w}_k$ is random and correlated with the $\hat{\beta_k}$, so the normal approximation given in (3) can be very crude. Indeed, it is well documented that confidence intervals based on (3) have very poor coverage for small and moderate K (see, e.g., Biggerstaff & Tweedie, 1997; Brockwell & Gordon, 2001). Various methods have been proposed to improve the confidence intervals (e.g., Hardy & Thompson, 1996; Biggerstaff & Tweedie, 1997; Brockwell & Gordon, 2001; Sidik & Jonkman, 2002; Henmi & Copas, 2010); however, none of these methods has been rigorously justified or widely accepted.

When individual participant data are available, one can maximize the loglikelihood

$$\sum_{k=1}^{K} \log \int_{\xi_k} \prod_{i=1}^{n_k} f_k(\mathscr{O}_{ki}; \beta + \xi_k, \eta_k)\,(2\pi\tau^2)^{-1/2} \exp(-\xi_k^2/2\tau^2)\,\mathrm{d}\xi_k. \quad (4)$$

Denote the resulting estimators of β, τ^2 and $(\eta_1,\ldots,\eta_K)$ by $\hat{\beta}_{\mathrm{ML}}$, $\hat{\tau}^2_{\mathrm{ML}}$ and $(\hat{\eta_1},\ldots,\hat{\eta_K})$. It is challenging to establish the theoretical properties of the maximum likelihood estimators, because the n_k are typically larger than K. This problem is similar to that encountered in the analysis of large clusters. The existing theory for random-effects models with large clusters requires that the variance component be known (Bellamy et al., 2005). The fact that the variance component needs to be estimated from the data poses major theoretical challenges.

In the present paper, we establish the asymptotic properties of the meta-analysis estimators $\hat{\beta_{\mathrm{MA}}}$ and $\hat{\tau}^2_{\mathrm{MA}}$ and the maximum likelihood estimators $\hat{\beta_{\mathrm{ML}}}$ and $\hat{\tau}^2_{\mathrm{ML}}$ for the situations of a fixed K and diverging K. We then investigate the asymptotic relative efficiency of $\hat{\beta_{\mathrm{MA}}}$ to

$\hat{\beta_{\mathrm{ML}}}$ and reveal that the former is at least as efficient as the latter. In addition, we develop a novel resampling technique that yields substantially more accurate inference than the normal approximation given in (3). Finally, we demonstrate the advantages of the new inference procedures by applying them to data from clinical trials on the treatment of cancer-associated anaemia.

2. Asymptotic theory

We assume that the n_k ($k = 1, \ldots, K$) are comparable and denote their median by n. We write $\tau^2 = \sigma^2/n$, where σ^2 is a constant, such that the between-study variability is of the order n^{-1} and thus comparable to the within-study variability $\hat{V_k}$. We estimate σ^2 by $\hat{\sigma}^2_{\mathrm{MA}} = n\,\hat{\tau}^2_{\mathrm{MA}}$ and $\hat{\sigma}^2_{\mathrm{ML}} = n\,\hat{\tau}^2_{\mathrm{ML}}$ in the meta-analysis and maximum likelihood estimation, respectively. Note that $\hat{\sigma}^2_{\mathrm{MA}}$ is a consistent method-of-moments estimator.

We assume the following regularity conditions.

Condition 1. The parameters (β, η_k, σ^2) lie in the interior of a compact set $\mathscr{C}_1 \times \mathscr{C}_{2k} \times \mathscr{C}_3$ within the parameter domain.

Condition 2. For (β_k, η_k) $\in \mathscr{C}_1 \times \mathscr{C}_{2k}$, the function $\log f_k(\mathscr{O}; \beta_k, \eta_k)$ is thrice continuously differentiable.

Condition 3. For $k = 1, \ldots, K$, $n_k = p_k n$ for some constant p_k within a compact interval in $(0, \infty)$.

Condition 4. The information matrix of $f_k(\mathscr{O}; \beta, \eta_k)$ is continuous in a neighbourhood of the true parameter value (β_0, η_{k0}), and its eigenvalues have positive lower and upper bounds uniformly for $k = 1, \ldots, K$.

In practice, K is small relative to n. Therefore, we investigate the asymptotic properties of the meta-analysis estimators and maximum likelihood estimators in the cases where K is fixed or diverges to ∞ at a slower rate than n. The results for the first case are stated below.

Theorem 1

Under Conditions 1–4,

$$\hat{\sigma}^2_{\mathrm{ML}} \to \mathscr{A}_{\mathrm{ML}} = \arg\max_{\sigma^2} \left\{ -\frac{1}{2} \sum_{k=1}^{K} \log(\mathscr{I}^{-1}_{\beta_k} + p_k \sigma^2) - \frac{1}{2} \sum_{k=1}^{K} \frac{\mathscr{L}^2_k}{v_k + p_k \sigma^2} + \frac{1}{2} \left(\sum_{k=1}^{K} \frac{p_k}{v_k + p_k \sigma^2} \right)^{-1} \left(\sum_{k=1}^{K} \frac{p_k^{1/2} \mathscr{L}_k}{v_k + p_k \sigma^2} \right)^2 \right\}$$

in distribution, where $\mathscr{I}_{\beta_k} = -E\{\partial^2 \log f_k(\mathscr{O}; \beta_0, \eta_{k0})/\partial \beta_k^2\}$, $\mathscr{L}_1, \ldots, \mathscr{L}_K$ *are independent zero-mean normal random variables with variances* $v_k + p_k \sigma_0^2$, v_k *is the limit of* $n_k \hat{V_k}$, *and* σ_0^2 *is the true value of* σ^2. *In addition,*

$$n^{1/2}(\hat{\beta}_{\mathrm{ML}}-\beta_0) \rightarrow \left(\sum_{k=1}^{K}\frac{p_k}{v_k+p_k\mathscr{A}_{\mathrm{ML}}}\right)^{-1}\sum_{k=1}^{K}\frac{p_k^{1/2}\mathscr{Z}_k}{v_k+p_k\mathscr{A}_{\mathrm{ML}}}$$

in distribution. Furthermore,

$$\hat{\sigma}_{\mathrm{MA}}^2 \rightarrow \mathscr{A}_{\mathrm{MA}}$$
$$= \max\left[0, \frac{\sum_{k=1}^{K}\tilde{v}_k^{-1}\{\mathscr{Z}_k/p_k^{1/2}-(\sum_{k=1}^{K}\tilde{v}_k^{-1})^{-1}\sum_{k=1}^{K}\tilde{v}_k^{-1}\mathscr{Z}_k/p_k^{1/2}\}^2-(K-1)}{\sum_{k=1}^{K}\tilde{v}_k^{-1}-\sum_{k=1}^{K}\tilde{v}_k^{-2}/\sum_{k=1}^{K}\tilde{v}_k^{-1}}\right]$$

in distribution, where $\tilde{v}_k=v_k/p_k$, and

$$n^{-1/2}(\hat{\beta}_{\mathrm{MA}}-\beta_0) \rightarrow \left(\sum_{k=1}^{K}\frac{p_k}{v_k+p_k\mathscr{A}_{\mathrm{MA}}}\right)^{-1}\sum_{k=1}^{K}\frac{p_k^{1/2}\mathscr{Z}_k}{v_k+p_k\mathscr{A}_{\mathrm{MA}}}$$

in distribution.

This theorem indicates that neither the maximum likelihood estimator $\hat{\beta}_{\mathrm{ML}}$ nor the metaanalysis estimator $\hat{\beta}_{\mathrm{MA}}$ is asymptotically normal. Their asymptotic distributions are mixtures of normal random variables, the mixing probabilities being random and correlated with the normal random variables. This phenomenon is different from the usual asymptotic theory (Cox & Hinkley, 1974, § 9.2) and is caused by the fact that K is fixed.

Before stating the asymptotic results for divergent K, we make two additional assumptions.

Condition 5. We have $n\rightarrow\infty$, $K \rightarrow \infty$, and $Kn^{-1/2}\rightarrow0$.

Condition 6. For any $\sigma^2 \in \mathscr{C}_3$, the following limit exists:

$$g(\sigma^2)=-\lim_{K\rightarrow\infty}\frac{1}{K}\sum_{k=1}^{K}\left\{\frac{v_k+p_k\sigma_0^2}{v_k+p_k\sigma^2}+\log\left(\mathscr{I}_{\beta_k}^{-1}+p_k\sigma^2\right)\right\},$$

where $g(\sigma^2)$ has a unique maximum at σ_{r}^2 in $\mathscr{C}_3$.

Remark 1

Condition 5 implies that the number of studies increases to infinity but not as fast as the study sizes. The function $g(\cdot)$ is the limit of the profile likelihood function for σ^2, so Condition 6 guarantees the convergence of $\hat{\sigma}_{\mathrm{ML}}^2$ to some unique value as $K \rightarrow \infty$.

Theorem 2

Under Conditions 1–6, $\hat{\sigma}_{\mathrm{ML}}^2 \rightarrow \sigma_{\mathrm{r}}^2$ *and* $\hat{\beta}_{\mathrm{ML}}\rightarrow\beta_0$ *in probability. In addition,*

$$(K\,n)^{1/2}(\hat{\beta}_{\mathrm{ML}}-\beta_0) \to N\{0,\ \sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\sum(\sigma_{\mathrm{r}}^2,\sigma_{\mathrm{r}}^2)\sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\}$$

in distribution, where

$$\sum(\sigma_1^2,\sigma_2^2)=\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{p_k(v_k+p_k\sigma_0^2)}{(v_k+p_k\sigma_1^2)(v_k+p_k\sigma_2^2)},$$

which is assumed to exist for any σ_1^2 and σ_2^2 in $\mathscr{C}_3$. Under Conditions 1–5, $\hat{\sigma}_{\mathrm{MA}}^2 \to \sigma_0^2$ and $\hat{\beta}_{\mathrm{MA}} \to \beta_0$ in probability, and $(K\,n)^{1/2}(\hat{\beta}_{\mathrm{MA}}-\beta_0) \to N\{0,\ \sum(\sigma_0^2,\sigma_0^2)^{-1}\}$ in distribution. Furthermore, $\mathrm{var}(\hat{\beta}_{\mathrm{ML}}) \quad \mathrm{var}(\hat{\beta}_{\mathrm{MA}})$ asymptotically, with equality if and only if $\sigma_{\mathrm{r}}^2=\sigma_0^2$ or $\tilde{v}_1 =\cdots=\tilde{v}_K$.

Theorem 2 provides some intriguing results. First, the maximum likelihood estimator for σ^2 may not even be consistent; it converges to σ_{r}^2 as defined in Condition 6. Second, the metaanalysis estimator $\hat{\beta}_{\mathrm{MA}}$ is at least as efficient as the maximum likelihood estimator $\hat{\beta}_{\mathrm{ML}}$. Both results contradict standard likelihood theory, which is applicable only to a large number of small clusters with fixed parameters. In our case, K is relatively small compared to n, and the variance component τ^2 changes with n.

3. A special case

To gain some insights into the asymptotic properties of the maximum likelihood estimators described in Theorems 1 and 2, we consider the special case of simple linear regression:

$$Y_{ki}=\alpha_k+(\beta+\xi_k)X_{ki}+\varepsilon_{ki} \quad (i=1,\ldots,n_k; k=1,\ldots,K),$$

where $\xi_k \sim N(0,\ \sigma^2/n)$, $\varepsilon_{ki} \sim N(0, 1)$, and ξ_k is independent of ε_{ki}. Define

$$m_{x1k}=\sum_{i=1}^{n_k}X_{ki}, \quad m_{x2k}=\sum_{i=1}^{n_k}X_{ki}^2, \quad m_{yk}=\sum_{i=1}^{n_k}Y_{ki}, \quad m_{yxk}=\sum_{i=1}^{n_k}Y_{ki}X_{ki}.$$

The loglikelihood function is, up to some constant,

$$-\frac{1}{2}\sum_{k=1}^{K}\log\left(\frac{n}{m_{x2k}}+\sigma^2\right)-\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_{ki}-\alpha_k-\beta X_{ki})^2+\frac{1}{2}\sum_{k=1}^{K}\frac{1}{n/\sigma^2+m_{x2k}}(m_{yxk}-\alpha_k m_{x1k}-\beta m_{x2k})^2.$$

After maximization over the α_k and β, we obtain the profile loglikelihood function for σ^2,

$$l_n(\sigma^2) = -\frac{1}{2}\sum_{k=1}^{K}\log\left(\frac{n}{m_{x2k}}+\sigma^2\right) - \frac{1}{2}\frac{n}{\sigma^2}\sum_{k=1}^{K}\frac{(n_k m_{yxk}-m_{yk}m_{x1k})^2/(n_k m_{x2k}-m_{x1k}^2)}{(n_k n/\sigma^2+n_k m_{x2k}-m_{x1k}^2)}$$
$$+\frac{1}{2}\frac{n}{\sigma^2}\frac{\{\sum_{k=1}^{K}(n_k m_{yxk}-m_{yk}m_{x1k})/(n_k^2/\sigma^2+n_k m_{x2k}-m_{x1k}^2)\}^2}{\sum_{k=1}^{K}(n_k m_{x2k}-m_{x1k}^2)/(n_k n/\sigma^2+n_k m_{x2k}-m_{x1k}^2)}.$$

Note that $\hat{\beta}_k = (n_k m_{yxk}-m_{yk}m_{x1k})/(n_k m_{x2k}-m_{x1k}^2)$ with variance v_k/n_k, where $v_k = n_k^2/(n_k m_{x2k}-m_{x1k}^2)$. We can rewrite the profile loglikelihood function as

$$l_n(\sigma^2) = -\frac{1}{2}\sum_{k=1}^{K}\log\left(\frac{n}{m_{x2k}}+\sigma^2\right) - \frac{n}{2}\sum_{k=1}^{K}\frac{(\hat{\beta}_k-\beta_0)^2}{\tilde{v}_k+\sigma^2} + \frac{n}{2}\frac{\left\{\sum_{k=1}^{K}(\hat{\beta}_k-\beta_0)/(\tilde{v}_k+\sigma^2)\right\}^2}{\sum_{k=1}^{K}1/(\tilde{v}_k+\sigma^2)}.$$

If K is fixed while n goes to ∞, then $\hat{\beta}_k = \beta_0 + (\sigma_0^2/n + v_k/n_k)^{1/2}Z_k$, where $Z_1,\dots,Z_K$ are independent standard normal. It follows that

$$l_n(\sigma^2) = -\frac{1}{2}\sum_{k=1}^{K}\log\left(\frac{n}{m_{x2k}}+\sigma^2\right) - \frac{1}{2}\sum_{k=1}^{K}\frac{(\tilde{v}_k+\sigma_0^2)Z_k^2}{\tilde{v}_k+\sigma^2} + \frac{1}{2}\frac{\left\{\sum_{k=1}^{K}(\tilde{v}_k+\sigma_0^2)^{1/2}Z_k/(\tilde{v}_k+\sigma^2)\right\}^2}{\sum_{k=1}^{K}1/(\tilde{v}_k+\sigma^2)},$$

which is the result for $\hat{\sigma}_{\mathrm{ML}}^2$ given in Theorem 1. When K diverges, Theorem 2 implies that the argument maximizing $K^{-1}l_n(\sigma^2)$, i.e., $\hat{\sigma}_{\mathrm{ML}}^2$, converges to σ_{r}^2, which maximizes the function

$$\lim_{K\to\infty}\frac{1}{K}\left[-\frac{1}{2}\sum_{k=1}^{K}\log\left\{\frac{1}{p_k E(X_{k1}^2)}+\sigma^2\right\} - \frac{1}{2}\sum_{k=1}^{K}\frac{\mathrm{var}(X_{k1})^{-1}/p_k+\sigma_0^2}{\mathrm{var}(X_{k1})^{-1}/p_k+\sigma^2}\right]. \quad (5)$$

Suppose that $p_k = 1$ for $k = 1,\dots, K$ and that the X_{ki} all have the same distribution with mean m_x and variance v_x. Then (5) becomes

$$-\frac{1}{2}\log\{(v_x+m_x^2)^{-1}+\sigma^2\} - \frac{1}{2}\frac{v_x^{-1}+\sigma_0^2}{v_x^{-1}+\sigma^2}.$$

Let σ_{m}^2 and σ_{M}^2 denote, respectively, the minimum and maximum of σ^2 in $\mathscr{C}_3$. It can be shown that if $\sigma_0^2 < 4m_x^2 v_x^{-1}(v_x+m_x^2)^{-1}-v_x^{-1}$, then $\sigma_{\mathrm{r}}^2 = \sigma_{\mathrm{m}}^2$; otherwise, σ_{r}^2 can be $\sigma_{\mathrm{m}}^2, \sigma_{\mathrm{M}}^2$ or a value that maximizes the above limit, and the maximizer has to be

$$-\frac{1}{2}(v_x^{-1}-\sigma_0^2) + \frac{1}{2}\left[(v_x^{-1}+\sigma_0^2)\{v_x^{-1}+\sigma_0^2-4m_x^2 v_x^{-1}(v_x+m_x^2)^{-1}\}\right]^{1/2},$$

which is σ_0^2 if $m_x = 0$. In particular, if $X_{ki} \sim 2 \times \text{Ber}(0.5) - 1 + m_x$, $\sigma_0^2 = 1$, $\sigma_m^2 = 0.001$ and $\sigma_M^2 = 2$, then the above derivation implies that $\sigma_r^2 = \sigma_m^2 = 0.001$ under $m_x = 2^{1/2}$ and $\sigma_r^2 = \{(1 - m_x^2)/(1 + m_x^2)\}^{1/2} = 0.6^{1/2}$ under $m_x = 0.5$. In both cases, $\sigma_r^2 \neq \sigma_0^2$; however, the asymptotic variances of $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_{\text{MA}}$ are the same since the $\tilde{v}_k$ ($k = 1, \dots, K$) are equal.

To induce unequal variances, we let $n_k = 50, 100, 400$ and 800 for every four studies, and let X_{ki} take the value 0 with probability a $\text{Un}(0.05, 0.5)$ random variable in each study and the value 2 otherwise. We set $a_k = 0$ ($k = 1, \dots, K$), $\beta_0 = 0.5$, $\sigma_0^2 = 1$, $\sigma_m^2 = 0$, $\sigma_M^2 = \infty$ and $n = 200$. Then the asymptotic relative efficiency of $\hat{\beta}_{\text{MA}}$ to $\hat{\beta}_{\text{ML}}$ is approximately 1·2 according to Theorem 2. Using Monte Carlo simulation, we found the empirical relative efficiencies to be approximately 1·037, 1·083, 1·125, 1·171 and 1·218 for $K = 100, 200, 300, 400$ and 500, respectively. The distributions of $\hat{\beta}_{\text{ML}}$, $\hat{\beta}_{\text{MA}}$, $\hat{\sigma}_{\text{ML}}^2$ and $\hat{\sigma}_{\text{MA}}^2$ for $K = 300$ are displayed in Fig. 1. The empirical means of $\hat{\beta}_{\text{ML}}$, $\hat{\beta}_{\text{MA}}$, $\hat{\sigma}_{\text{ML}}^2$ and $\hat{\sigma}_{\text{MA}}^2$ are 0·50, 0·50, 0·00076 and 1·00, respectively, and the corresponding standard errors are 0·0073, 0·0069, 0·0263 and 0·169.

4. Inference procedures

When K is large, we make inference about β_0 by using the asymptotic normality of $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_{\text{MA}}$ described in Theorem 2. The asymptotic variance of $(nK)^{1/2}(\hat{\beta}_{\text{ML}} - \beta_0)$ can be consistently estimated by $\hat{\sum}(\hat{\sigma}_{\text{ML}}^2, \hat{\sigma}_{\text{MA}}^2)^{-1} \hat{\sum}(\hat{\sigma}_{\text{ML}}^2, \hat{\sigma}_{\text{ML}}^2) \hat{\sum}(\hat{\sigma}_{\text{ML}}^2, \hat{\sigma}_{\text{MA}}^2)^{-1}$, where

$$\hat{\sum}(\sigma_1^2, \sigma_2^2) = \frac{1}{K} \sum_{k=1}^{K} \frac{p_k(n_k \hat{V}_k + p_k \hat{\sigma}_{\text{MA}}^2)}{(n_k \hat{V}_k + p_k \sigma_1^2)(n_k \hat{V}_k + p_k \sigma_2^2)}.$$

The asymptotic variance of $(nK)^{1/2}(\hat{\beta}_{\text{MA}} - \beta_0)$ can be consistently estimated by $\hat{\sum}(\hat{\sigma}_{\text{MA}}^2, \hat{\sigma}_{\text{MA}}^2)^{-1}$. The normal approximation based on this variance estimator is equivalent to the normal approximation given in (3), which is the DerSimonian–Laird method.

When K is small, the normal approximations to the distributions of $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_{\text{MA}}$ are no longer accurate, so we appeal to Theorem 1. Without knowledge of σ_0^2, however, it is not possible to estimate the asymptotic distributions given in Theorem 1. To deal with this problem, we propose a double resampling approach to meta-analysis. Theorem 1 shows that, for fixed K, the asymptotic distribution of $n^{1/2}(\hat{\beta}_{\text{MA}} - \beta_0)$ depends on the asymptotic distribution of $\hat{\sigma}_{\text{MA}}^2$. Thus, we first simulate σ^2 from the distribution of $\mathscr{A}_{\text{MA}}$, in which $\mathscr{Z}_k \sim N(0, n_k \hat{V}_k + p_k \hat{\sigma}_{\text{MA}}^2)$. For each sampled σ^2, say σ_s^2, we then generate β as

$$\hat{\beta}_{\text{MA}} + n^{-1/2} \left(\sum_{k=1}^{K} \frac{p_k}{n_k \hat{V}_k + p_k \sigma_s^2} \right)^{-1} \sum_{k=1}^{K} \frac{p_k^{1/2} \mathscr{Z}_k}{n_k \hat{V}_k + p_k \sigma_s^2},$$

where $\mathscr{Z}_k \sim N(0, \; n_k \hat{V}_k + p_k \sigma_s^2)$. Operationally, this resampling procedure is equivalent to the following: simulate τ^2 from equation (2), in which $\hat{\beta}_k \sim N(\hat{\beta}_{MA}, \; \hat{V}_k + \hat{\tau}_{MA}^2)$; for each sampled τ^2, say τ_s^2, simulate β from equation (1), in which $\hat{\beta}_k \sim N(\hat{\beta}_{MA}, \; \hat{V}_k + \tau_s^2)$ and $\hat{w}_i = (\hat{V}_k + \tau_s^2)^{-1}$. We repeat this process B times to obtain B^2 values of β. The empirical distribution of those values is used to make inference about β.

5. Simulation studies

We used simulation to evaluate the proposed inference procedures. To facilitate comparisons with existing methods, we adopted the simulation set-up of Brockwell & Gordon (2001), which corresponds to a typical scenario of estimating a log odds ratio. Specifically, we generated the parameter estimates as

$$\hat{\beta}_k = \beta + N(0, \; \tau^2 + \hat{V}_k) \quad (k = 1, \ldots, K),$$

where $\beta = 0\cdot 5$ and the $\hat{V}_k$ $(k = 1, \ldots, K)$ are realizations from a χ_1^2 distribution multiplied by $0\cdot 25$ and then restricted to lie within the interval $(0\cdot 009, \; 0\cdot 6)$. We varied τ^2 from 0 to $0\cdot 1$ and K from 5 to 50. The corresponding values of I^2 (Higgins et al., 2003) are displayed in the Supplementary Material. For each simulated dataset, we obtained the 95% confidence intervals for β using the new resampling approach with $B = 1000$, the DerSimonian–Laird method, the profile likelihood method of Hardy & Thompson (1996), and the resampling method of Jackson & Bowden (2009). The coverage probabilities based on 10 000 replicates are shown in Fig. 2 and the Supplementary Material; the corresponding mean widths are shown in the Supplementary Material.

The new method has reasonable coverage probabilities, especially when τ^2 is small or K is large. Its coverage probabilities are always higher than those of the DerSimonian–Laird method. The differences between the two methods become smaller as K increases. Indeed, both methods have correct coverage probabilities when K is 50. The Jackson–Bowden method is very conservative. The new method has better coverage than the profile likelihood method when K is small and τ^2 is not too small. By comparing our Fig. 1 with Fig. 4 of Brockwell & Gordon (2001), Fig. 1 of Sidik & Jonkman (2002), and Fig. 1 of Henmi & Copas (2010), we see that the new method outperforms the other methods. The widths of the confidence intervals are similar for the new and profile likelihood methods, which are bigger than those of the DerSimonian–Laird method.

6. Real-data example

The erythropoiesis-stimulating agents erythropoietin and darbepoetin are approved to treat chemotherapy-associated anaemia in patients with nonmyeloid malignancies. To evaluate mortality rates associated with the administration of these agents for the treatment of anaemia in cancer patients, Bennett et al. (2008) conducted a systematic review of 52 phase III clinical trials with 13 611 patients that compared the erythropoiesis-stimulating agents with placebo or standard care with respect to mortality. The estimated hazard ratios and 95%

confidence intervals are shown in Fig. 2 of Bennett et al. (2008). Using the DerSimonian–Laird method, Bennett et al. (2008) obtained a hazard ratio estimate of 1·10 with a 95% confidence interval of (1·01, 1·20), which raised concerns about the safety of administering these agents to patients with cancer. Our resampling method yields a 95% confidence interval of (1·006, 1·201), which is very close to the DerSimonian–Laird counterpart. The corresponding intervals are (0·973, 1·242) and (0·986, 1·230) for the Jackson–Bowden and Hardy–Thompson methods, respectively.

Bennett et al. (2008) also applied the DerSimonian–Laird method to a subset of six trials consisting of 2089 patients who did not receive chemotherapy or radiation therapy and obtained a hazard ratio estimate of 1·29 with a 95% confidence interval of (1·00, 1·67). Our resampling method yields a 95% confidence interval of (0·851, 1·963), which is considerably wider. In this case, $\hat{\tau}^2_{\mathrm{MA}} = 0$, so the DerSimonian–Laird method is actually the same as the fixed-effects method. By contrast, our resampling approach accounts for the variation in the estimation of τ^2 and thus will not reduce to the fixed-effects method even when the point estimate of τ^2 is zero. The Jackson–Bowden and Hardy–Thompson methods yield intervals of (0·798, 2·094) and (0·703, 1·792), respectively.

7. Remarks

Effect sizes tend to vary among study populations because of differences in demographic and environmental factors. Furthermore, the treatments or outcomes may not be identical across clinical trials, so the treatment effects may differ even for similar patient populations. In genetic association studies, different definitions and measurements of phenotypes, as well as different collections and manipulations of genotype data, also contribute to between-study heterogeneity. Thus, it is important to allow for heterogeneity through the use of random-effects models, especially when one is interested in parameter estimation rather than hypothesis testing. The confidence intervals under fixed-effects models have extremely poor coverage under even mild heterogeneity (e.g., Brockwell & Gordon, 2001; Henmi & Copas, 2010).

The prevailing approach to random-effects meta-analysis is the DerSimonian–Laird method. Indeed, DerSimonian & Laird's 1986 paper has been cited more than 10 000 times in the Web of Science database. Our paper provides a rigorous asymptotic theory for the DerSimonian–Laird estimator in cases where the number of studies is fixed or divergent. In addition, we propose a resampling technique that yields more accurate inference than the commonly adopted normal approximation.

In most applications, the number of studies is much smaller than the study sizes, so that Condition 5 holds. It would be interesting to consider situations in which K diverges at the same rate as n or n^r for some $r \geq 1/2$. Such an extension would require careful examination of the higher-order expansion in the quadratic approximation to the profile loglikelihood of each study.

The random-effects model proposed by DerSimonian & Laird (1986) and considered in this paper assumes that the random effect is normally distributed. Additional simulation studies

have revealed that our resampling procedure performs satisfactorily under other random-effects distributions; see Fig. 3. Recently, Wang et al. (2010) proposed a nonparametric inference procedure for the percentiles of the random-effects distribution. Their method works well for small K but may not be statistically efficient.

For ethical and logistical reasons, individual participant data are not as easily accessible as summary statistics. Our work shows that it is not necessary to collect individual participant data. In fact, maximum likelihood analysis of individual participant data can give worse results than meta-analysis of summary statistics: the maximum likelihood estimator of σ^2 may not be consistent as $K \to \infty$, and the maximum likelihood estimator of β is always less efficient than the meta-analysis estimator.

We have assumed that β is a scalar; however, all our results can be extended to multivariate random-effects models (e.g., Jackson et al., 2010; Chen et al., 2012). Specifically, the meta-analysis estimator of β still takes the form of (1), whereas expression (2) is replaced by an estimator of the covariance matrix of the random effects (e.g., Chen et al., 2012). The basic conclusions of Theorems 1 and 2 continue to hold; a key change in the proofs is to approximate the profile likelihood function by a quadratic function based on the multivariate version of the Laplace approximation. Our resampling procedure remains the same except that multivariate versions of the estimators are used.

Our work can be applied to the analysis of large clusters in other contexts. The most rigorous theory for the analysis of large clusters was provided by Bellamy et al. (2005), who quantified the bias of the penalized quasilikelihood estimator of the cluster-level covariate effect in generalized linear mixed models for group-randomized trials under the assumption of a known variance component. Our framework covers this case with an unknown variance component upon setting $\beta_k = X_k \gamma + \xi_k$ $(k = 1, \ldots, K)$, where X_k is a group-level treatment indicator. More generally, we may assume that $\beta_k = X_k \gamma_k$ $(k = 1, \ldots, K)$, where $\gamma_k = \gamma + \xi_k$ and X_k is a cluster-level covariate. Our theory suggests that meta-analysis of summary statistics is preferable to maximum likelihood analysis of individual participant data in such situations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bellamy SL, Li Y, Lin X, Ryan LM. Quantifying PQL bias in estimating cluster-level covariate effects in generalized linear mixed models for group-randomized trials. Statist Sinica. 2005; 15:1015–32.

Bennett CL, Silver SM, Djulbegovic B, Samaras AT, Blau CA, Gleason KJ, Barnato SE, Elverman KM, Courtney DM, McKoy JM, Edwards BJ, Tigue CC, Raisch DW, Yarnold PR, Dorr DA, Kuzel TM, Tallman MS, Trifilio SM, West DP, Lai SY, et al. Venous thromboembolism and mortality

associated with recombinant erythropoietin and darbepoetin administration for the treatment of cancer-associated anemia. J Am Med Assoc. 2008; 299:914–24.

Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Statist Med. 1997; 16:753–68.

Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Statist Med. 2001; 20:825–40.

Chalmers I, Sandercrock P, Wennberg J. The Cochrane collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann New York Acad Sci. 1993; 703:156–65. [PubMed: 8192293]

Chen H, Manning AK, Dupuis J. A method of moments estimator for random effect multivariate meta-analysis. Biometrics. 2012; 68:1278–84. [PubMed: 22551393]

Cox, DR.; Hinkley, DV. Theoretical Statistics. New York: Chapman & Hall; 1974.

DerSimonian R, Laird N. Meta-analysis in clinical trials. Contr Clin Trials. 1986; 7:177–88.

Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nature Rev Genet. 2013; 14:379–89. [PubMed: 23657481]

Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. Statist Med. 1996; 15:619–29.

Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. Statist Med. 2010; 29:2969–83.

Higgins J, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J. 2003; 327:557–60. [PubMed: 12958120]

Jackson D, Bowden J. A re-evaulation of the 'quantile approximation method' for random effects meta-analysis. Statist Med. 2009; 28:338–48.

Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Statist Med. 2010; 29:1282–97.

Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual level data in meta-analysis. Biometrika. 2010; 97:321–32. [PubMed: 23049122]

Lin DY, Zeng D, Tang ZZ. Quantitative trait analysis in sequencing studies with trait-dependent sampling. Proc Nat Acad Sci. 2013; 110:12247–52. [PubMed: 23847208]

Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. Biometrics. 1999; 55:1221–3. [PubMed: 11315071]

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nature Rev Genet. 2008; 9:356–69. [PubMed: 18398418]

Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics. 1998; 54:317–22. [PubMed: 9544524]

Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. Molec Psychiatry. 2009; 14:10–17. [PubMed: 19002139]

Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. Statist Med. 2002; 21:3153–9.

Sutton AJ, Higgins JPT. Recent developments in meta-analysis. Statist Med. 2008; 27:625–50.

Sutton, AJ.; Abrams, KR.; Jones, DR.; Sheldon, TA.; Song, F. Methods for Meta-Analysis in Medical Research. Chichester: Wiley; 2000.

van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. New York: Springer; 1996.

Wang R, Tian L, Cai T, Wei LJ. Nonparametric inference procedure for percentiles of the random effects distribution in meta-analysis. Ann Appl Statist. 2010; 4:520–32.

Whitehead, A. Meta-Analysis of Controlled Clinical Trials. Chichester: Wiley; 2002.

Appendix. Proofs of Theorems 1 and 2

The proofs of both theorems rely on a local expansion of the loglikelihood function around the maximum likelihood estimators. By Condition 2 and Theorem 2.83 of van der Vaart & Wellner (1996), the class of functions $\{\log f_k(\mathcal{O}; \beta_k, \eta_k) : (\beta_k, \eta_k) \in \mathscr{C}_1 \times \mathscr{C}_{2k}\}$ and the corresponding classes of first and second derivatives are uniformly Donsker and Glivenko–Cantelli with respect to any probability measure. We will use this fact repeatedly. Let $\wp_{nk}$ denote the empirical measure for $(\mathcal{O}_{k1}, \ldots, \mathcal{O}_{k, n_k})$ conditional on ξ_k, and let $\wp_k$ be the conditional expectation under the density $f_k(\cdot; \beta_{k0}, \eta_{k0})$, where β_{k0} is the true value of β_k. Then the kth summand in (4) can be written as

$$l_k(\beta, \sigma^2, \eta_k) = -\frac{1}{2} \log \frac{2\pi\sigma^2}{n} + \log \int_{\xi_k} \exp\left(-n_k \left[\mathscr{P}_{nk}\{-\log f_k(\mathcal{O}; \beta + \xi_k, \eta_k)\} + \frac{n\xi_k^2}{2n_k\sigma^2}\right]\right) \mathrm{d}\xi_k.$$

Consider an open set $\mathcal{N}$ for $(\beta, \eta_1, \ldots, \eta_K, \sigma^2)$ with $|\beta - \beta_0| \quad M(nK)^{-1/2}$, $\|\eta_k - \eta_{k0}\| \quad Mn^{-1/2}$ ($k = 1, \ldots, K$) for some large M to be chosen later, and $\sigma^2 \in \mathscr{C}_3$. We will show that with probability tending to 1, there exists a local maximizer of $\sum_{k=1}^{K} l_k(\beta, \sigma^2, \eta_k)$ in this neighbourhood. The proof of existence consists of four main steps. First, we obtain a Laplace approximation to the integral in $l_k(\beta, \sigma^2, \eta_k)$. Second, we show that for each (β, σ^2) there exist estimators $\hat{\eta}_k(\beta, \sigma^2)$ ($k = 1, \ldots, K$) which maximize $\sum_{k=1}^{K} l_k(\beta, \sigma^2, \eta_k)$, and we obtain the profile loglikelihood function $\sum_{k=1}^{K} \mathrm{pl}_k(\beta, \sigma^2)$, where $\mathrm{pl}_k(\beta, \sigma^2) = l_k\{\beta, \sigma^2, \hat{\eta}_k(\beta, \sigma^2)\}$. Third, we show that there exists an estimator $\hat{\beta}_{\mathrm{ML}}(\sigma^2)$ which maximizes $\sum_{k=1}^{K} \mathrm{pl}_k(\beta, \sigma^2)$ for each σ^2. Finally, we show that there exists an estimator for σ^2, denoted by $\hat{\sigma}_{\mathrm{ML}}^2$, which maximizes $\sum_{k=1}^{K} \mathrm{pl}_k\{\hat{\beta}_{\mathrm{ML}}(\sigma^2), \sigma^2\}$, so we obtain $\hat{\beta}_{\mathrm{ML}} = \hat{\beta}_{\mathrm{ML}}(\hat{\sigma}_{\mathrm{ML}}^2)$ and $\hat{\eta}_k = \hat{\eta}_k(\hat{\beta}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}}^2)$. These four steps are detailed in the Supplementary Material.

Proof of Theorem 1

When K is fixed, $\hat{\sigma}_{\mathrm{ML}}^2$ does not converge to any constant. It follows from equation (S5) in the Supplementary Material and the arg max continuous mapping theorem (van der Vaart & Wellner, 1996, Theorem 3.2.2) that

$$\hat{\sigma}_{\mathrm{ML}}^2 \to \arg\max \left\{-\frac{1}{2}\sum_{k=1}^{K} \log(\mathscr{I}_{\beta_k}^{-1} + p_k\sigma^2) - \frac{1}{2}\sum_{k=1}^{K} \frac{\mathscr{L}_k^2}{v_k + p_k\sigma^2} + \frac{1}{2}\left(\sum_{k=1}^{K} \frac{p_k}{v_k + p_k\sigma^2}\right)^{-1} \left(\sum_{k=1}^{K} \frac{p_k^{1/2}\mathscr{L}_k}{v_k + p_k\sigma^2}\right)^2\right\}$$

in distribution. It then follows from equation (S4) in the Supplementary Material that

$$n^{1/2}(\hat{\beta}_{\mathrm{ML}}-\beta_0) \to \left(\sum_{k=1}^{K}\frac{p_k}{v_k+p_k\mathscr{A}_{\mathrm{ML}}}\right)^{-1}\sum_{k=1}^{K}\frac{p_k^{1/2}\mathscr{Z}_k}{v_k+p_k\mathscr{A}_{\mathrm{ML}}}$$

in distribution. Likewise, the asymptotic distribution for $(\hat{\sigma}_{\mathrm{MA}}^2, \hat{\beta_{\mathrm{MA}}})$ can be obtained from the arg max continuous mapping theorem.

Proof of Theorem 2

We examine the random process $Q_n(\sigma^2)$ given in (S5) of the Supplementary Material. Let

$$\mathscr{Z}_{nk}(\sigma^2)=\frac{n_k\{\xi_{k0}+v_k(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}}{n^{1/2}(v_k+p_k\sigma^2)}.$$

Then

$$Q_n(\sigma^2)=-\frac{1}{2}\sum_{k=1}^{K}\log\left(p_k\sigma^2\mathscr{I}_{\beta_k}+1\right)-\frac{1}{2}\sum_{k=1}^{K}\frac{n}{n_k}\mathscr{Z}_{nk}^2(\sigma^2)(v_k+p_k\sigma^2)+\frac{n}{2}\left(\sum_{k=1}^{K}\frac{n_k}{v_k+p_k\sigma^2}\right)^{-1}\left\{\sum_{k=1}^{K}\mathscr{Z}_{nk}(\sigma^2)\right\}^2.$$

Consider the random process $K^{-1/2}\sum_{k=1}^{K}\mathscr{Z}_{nk}(\sigma^2)$ ndexed by σ^2. We verify the conditions in Theorem 2.11.1 of van der Vaart & Wellner (1996). First, we verify the Lindeberg condition,

$$\sum_{k=1}^{K}K^{-1}E\left[\max_{\sigma^2\in(\sigma_{\mathrm{m}}^2,\sigma_{\mathrm{M}}^2)}\mathscr{Z}_{nk}^2(\sigma^2)I\left\{K^{-1/2}\max_{\sigma^2\in(\sigma_{\mathrm{m}}^2,\sigma_{\mathrm{M}}^2)}|\mathscr{Z}_{nk}(\sigma^2)|>\delta\right\}\right]\to 0, \quad \delta>0.$$

By the Markov inequality, the left-hand side is bounded by $\sum_{k}E\{\mathscr{Z}_{nk}^4(\sigma_{\mathrm{m}}^2)\}/(\sigma K)^2$. In addition,

$$
\begin{aligned}
E(\xi_{k0}^4) &=2\sigma^4/n^2,\\
E\{|(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)|^4\} &=n_k^{-4}\{E(|S_k^*|^4)n_k+n_k(n_k-1)v_k\}\{1+O(n^{-1})\},\\
E\left[\xi_{k0}^2\{(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}^2\right] &=E\left(\xi_{k0}^2 E\left[\{(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}^2|\xi_{k0}\right]\right)\\
&=E\{\xi_{k0}^2\mathrm{var}(S_k^*|\xi_{k0})/n_k\}\\
&=v_k\sigma^2(nn_k)^{-1}\{1+O(n^{-1})\},\\
E\{\xi_{k0}^3(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\} &=E\left[\xi_{k0}^3 E\{(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)|\xi_{k0}\}\right]=0,\\
\left|E\left[\xi_{k0}\{(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}^3\right]\right| &=\left|E\left(\xi_{k0}E\left[\{(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}^3|\xi_{k0}\right]\right)\right|\\
&\le E\{|\xi_{k0}|n_k^{-2}E(|S_k^*|^3|\xi_{k0})\}\le c_3 n^{-5/2}
\end{aligned}
$$

for some constant c_3. Thus, $\sum_k E\{\mathscr{Z}_{nk}^4(\sigma_m^2)\}/(\delta K)^2 \le c_4 K^{-1} \to 0$, where c_4 is a constant. For the second condition in Theorem 2.11.1 of van der Vaart & Wellner (1996), we note that, by the mean value theorem,

$$K^{-1}\sum_{k=1}^K E\left[\{\mathscr{Z}_{nk}(\sigma_1^2)-\mathscr{Z}_{nk}(\sigma_2^2)\}^2\right] \le |\sigma_1^2-\sigma_2^2|^2\sum_{k=1}^K E\left[\frac{p_k^2 n_k\{\xi_{k0}+v_k(\mathscr{P}_{nk}-\mathscr{P})(S_k^*)\}^2}{(v_k+p_k\sigma_m^2)^2}\right] \to 0$$

as $|\sigma_1^2-\sigma_2^2| \to 0$. Since $\mathscr{C}_3$ is one-dimensional, it is easy to see that the random entropy condition in Theorem 2.11.1 holds. In light of Condition 6, Theorem 2.11.1 implies that $K^{-1/2}\sum_{k=1}^K \mathscr{Z}_{nk}(\sigma^2)$ converges weakly to a zero-mean Gaussian process with covariance function $\sum(\sigma_1^2,\sigma_2^2)$ between σ_1^2 and σ_2^2. Thus, the third term of $Q_n(\sigma^2)$ is $o_p(K)$ uniformly in σ^2. Similarly, for the second term of $Q_n(\sigma^2)$,

$$-\frac{1}{2K}\sum_{k=1}^K \frac{n}{n_k}\mathscr{Z}_{nk}^2(\sigma^2)(v_k+p_k\sigma^2) \to -\frac{1}{2}\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^K \frac{v_k+p_k\sigma_0^2}{v_k+p_k\sigma^2}$$

in probability. Combining the above results, we conclude that, in probability,

$$\frac{1}{K}\left\{Q_n(\sigma^2)+\frac{1}{2}\sum_{k=1}^K \log(\mathscr{I}_{\beta_k})\right\} \to \frac{1}{2}g(\sigma^2),$$

which has a unique maximum at σ_r^2 according to Condition 6.

We are ready to show that $\hat{\sigma}_{ML}^2 \to \sigma_r^2$ in probability. Since $\hat{\sigma}_{ML}^2$ is bounded, we can always choose a further subsequence from any subsequence, still denoted by $\hat{\sigma}_{ML}^2$, such that $\hat{\sigma}_{ML}^2 \to \tilde{\sigma}^2$. Because $\sum_{k=1}^K \text{pl}_k\{\hat{\beta}_{ML}(\hat{\sigma}_{ML}^2),\hat{\sigma}_{ML}^2\} \ge \sum_{k=1}^K \text{pl}_k\{\hat{\beta}_{ML}(\sigma_r^2),\sigma_r^2\}$, it follows from equation (S5) in the Supplementary Material that

$$\frac{1}{K}\left\{Q_n(\hat{\sigma}_{ML}^2)+\frac{1}{2}\sum_{k=1}^K \log(\mathscr{I}_{\beta_k})\right\} \ge \frac{1}{K}\left\{Q_n(\sigma_r^2)+\frac{1}{K}\sum_{k=1}^K \log(\mathscr{I}_{\beta_k})\right\} -O_p(1)n^{-1/2}.$$

Taking the limit on both sides yields $g(\tilde{\sigma}^2) \ge g(\sigma_r^2)$. Therefore $\tilde{\sigma}^2=\sigma_r^2$.

In light of (S4) in the Supplementary Material, for the local maximum likelihood estimator $\hat{\beta}_{ML}=\hat{\beta}_{ML}(\hat{\sigma}_{ML}^2)$, we have

$$(Kn)^{1/2}(\hat{\beta}_{\mathrm{ML}}-\beta_0)=(Kn)^{1/2}\left(\sum_{k=1}^{K}\frac{n_k}{v_k+p_k\sigma^2}\right)^{-1}\left[\sum_{k=1}^{K}\frac{n_k}{v_k+p_k\sigma^2}\{\xi_{k0}+v_k(\mathscr{P}_{nk}-\mathscr{P}_k)(S_k^*)\}\right]+o_{\mathrm{p}}(1)$$

$$=\left(\frac{1}{nK}\sum_{k=1}^{K}\frac{n_k}{v_k+p_k\hat{\sigma}_{\mathrm{ML}}^2}\right)^{-1}\left\{\sum_{k=1}^{K}K^{-1/2}\mathscr{Z}_{nk}(\hat{\sigma}_{\mathrm{ML}}^2)\right\}+o_{\mathrm{p}}(1).$$

By Theorem 2.11.1 of van der Vaart & Wellner (1996) and the convergence of $\hat{\sigma}_{\mathrm{ML}}^2$ to σ_{r}^2,

$$(Kn)^{1/2}(\hat{\beta}_{\mathrm{ML}}-\beta_0)\to N\{0,\ \sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\sum(\sigma_{\mathrm{r}}^2,\sigma_{\mathrm{r}}^2)\sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\}$$

in distribution.

We now derive the asymptotic distribution of $\hat{\beta_{\mathrm{MA}}}$. Because $\hat{\sigma}_{\mathrm{MA}}^2\to\sigma_0^2$ in probability as $K\to\infty$, we have $\sup_k|\hat{w}_k/n-w_k|=o_{\mathrm{p}}(1)$, where $w_k=p_k(v_k+p_k\sigma_0^2)^{-1}$. Thus,

$$(Kn)^{1/2}(\hat{\beta}_{\mathrm{MA}}-\beta_0)=\frac{\sum_{k=1}^{K}K^{1/2}(\hat{w}_k/n)n^{1/2}(\hat{\beta}_k-\beta_0)}{\sum_{k=1}^{K}(\hat{w}_k/n)}$$

$$=\frac{\sum_{k=1}^{K}K^{1/2}w_kn^{1/2}(\hat{\beta}_k-\beta_0)}{\sum_{k=1}^{K}w_k}\{1+o_{\mathrm{p}}(1)\}.$$

Since $\max_k w_k/\Sigma_k w_k\to 0$ by Condition 4, the Lindeberg–Feller central limit theorem yields that, in distribution,

$$\frac{\sum_{k=1}^{K}K^{1/2}w_kn^{1/2}(\hat{\beta}_k-\beta_0)}{\sum_{k=1}^{K}w_k}\to N\{0,\ \sum(\sigma_0^2,\sigma_0^2)^{-1}\}.$$

By the Cauchy–Schwarz inequality,

$\sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\sum(\sigma_{\mathrm{r}}^2,\sigma_{\mathrm{r}}^2)\sum(\sigma_{\mathrm{r}}^2,\sigma_0^2)^{-1}\geq\sum(\sigma_0^2,\sigma_0^2)^{-1}$, and equality holds if and only if $(\tilde{v}_k+\sigma_{\mathrm{r}}^2)$ is proportional to $(\tilde{v}_k+\sigma_0^2)$ for $k=1,\dots,K$. This condition is met if $\sigma_{\mathrm{r}}^2=\sigma_0^2$ or the $\tilde{v}_k$ ($k=1,\dots,K$) are all equal. The lower bound is the asymptotic variance of $\hat{\beta_{\mathrm{MA}}}$. Thus $\hat{\beta_{\mathrm{MA}}}$ has a smaller asymptotic variance than $\hat{\beta}_{\mathrm{ML}}$, unless $\sigma_{\mathrm{r}}^2=\sigma_0^2$ or the $\tilde{v}_k$ are all the same.
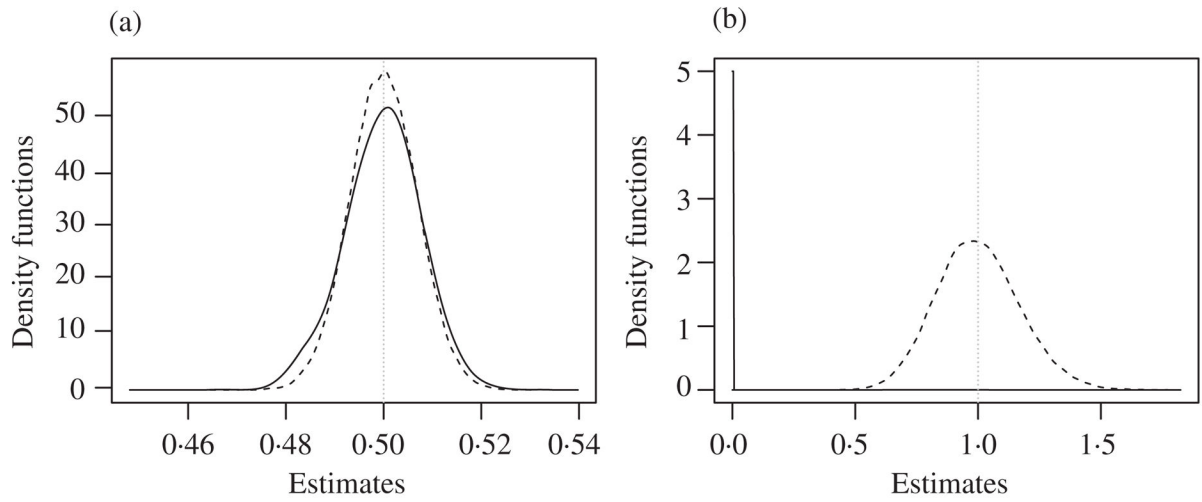
Fig. 1.

Estimated density functions of: (a) $\hat{\beta}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{MA}}$; (b) $\hat{\sigma}^2_{\mathrm{ML}}$ and $\hat{\sigma}^2_{\mathrm{MA}}$ in simple linear regression. In each panel, the solid curve corresponds to the maximum likelihood estimator and the dashed curve to the meta-analysis estimator.
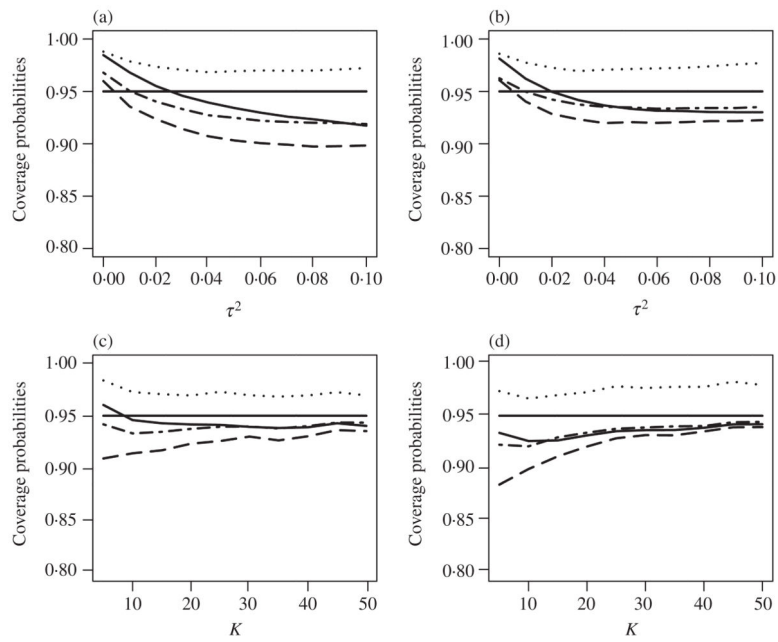
Fig. 2.
Empirical coverage probabilities of nominal 95% confidence intervals plotted against τ^2 for
(a) $K = 10$ and (b) $K = 20$, and plotted against K for (c) $\tau^2 = 0.03$ and (d) $\tau^2 = 0.07$. In each
panel, the different curves correspond to the new resampling method (solid), the
DerSimonian–Laird method (dashed), the Jackson–Bowden method (dotted), and the
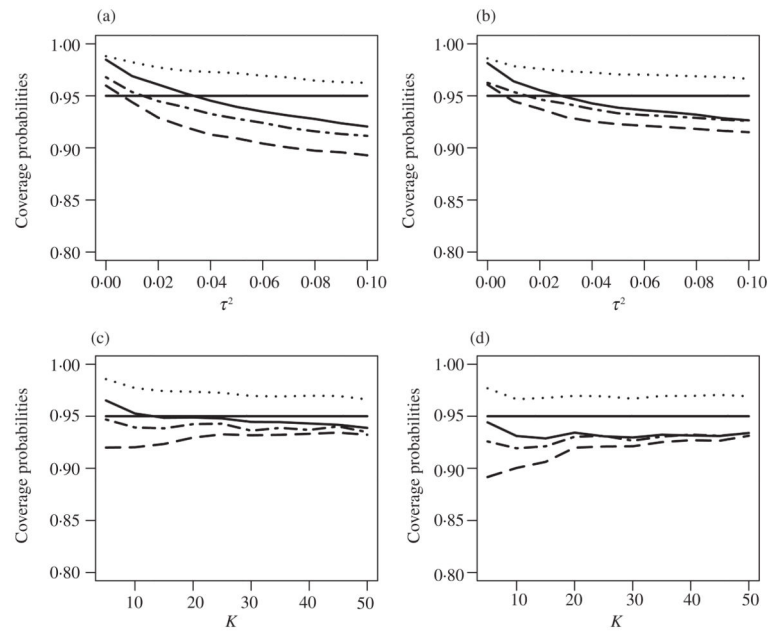Hardy–Thompson method (dot-dash).

Fig. 3.
Empirical coverage probabilities of nominal 95% confidence intervals when the random effects are from the $\tau \times$ Ga(1, 1) distribution centred at its mean, plotted against τ^2 for (a) K = 10 and (b) $K = 20$, and plotted against K for (c) $\tau^2 = 0.03$ and (d) $\tau^2 = 0.07$. In each panel, the different curves correspond to the new resampling method (solid), the DerSimonian–Laird method (dashed), the Jackson–Bowden method (dotted), and the Hardy–Thompson method (dot-dash).