On Estimating Residual Heterogeneity in Random-Effects Meta-Regression: A Comparative Study

Thammarat Panityakul [1] [*], Chinnaphong Bumrungsup [2] , Guido Knapp [3]

[1] *Department of Mathematics and Statistics, Thammasat University*
Pathumthani 12121, Thailand
E-mail: thammarat.thammasat@gmail.com

[2] *Department of Mathematics and Statistics, Thammasat University*
Pathumthani 12121, Thailand
E-mail: chinnaphong@msn.com

[3] *Department of Statistics, TU Dortmund University*
44221 Dortmund, Germany
E-mail: guido.knapp@tu-dortmund.de

Abstract

We consider six different estimators of residual heterogeneity in random-effects meta-regression, five estimators already known and implemented in the R package `metafor` and one estimator not yet considered in random-effects meta-regression. In a numerical study, we investigate the properties of these residual heterogeneity estimators as well as the impact of these estimators on the properties of the regression parameter estimates. It turns out that the new estimator performs quite well in terms of bias and mean squared error. The impact of the different residual heterogeneity estimators on the actual confidence coefficient of confidence intervals for regression parameters can be substantially different as shown in the numerical study.

Keywords: Binomial-normal hierarchical model; covariate; meta-analysis; relative risk.

1. Introduction

Meta-analysis aims to compare and possibly combine estimates of effect across related studies. For example, in a meta-analysis of clinical trials, the overall effect of a treatment can be expressed as standardized difference of means for normal responses or as relative risk or odds ratio for binary responses. Methods for providing such an overall estimate are well known[1,2]. Methods which incorporates a between-study component of variance for the treatment effect are based on random-effects models; the between-study variance represents the excess variation in observed treatment effects over that expected from the imprecision of results within each study. Between-study variance is also often called heterogeneity.

When substantial heterogeneity is present, un-

---

[*]Permanent address: Department of Mathematics and Statistics, Prince of Songkla University, Hat Yai Campus, Songkla 90112, Thailand.

derstanding the possible causes of heterogeneity is crucial. One possibility is to incorporate study-specific covariates in the random-effects model that then leads to the notion meta-regression. In random-effects meta-regression inference, that is, in a meta-regression model allowing for residual heterogeneity, accurate estimation of the variances of the parameter estimates is fundamental, as always in statistical inference. Accurate estimation of the residual heterogeneity is a first step in this process as the reduction of residual heterogeneity compared to meta-analysis heterogeneity can be used to interpret whether the study-specific covariates may or may not explain heterogeneity.

Several estimators of heterogeneity or residual heterogeneity have been already proposed. We consider five estimators already implemented in the R package `metafor` (Ref. 3), namely Hedges (HE) estimator, DerSimonian-Laird (DL) estimator, maximum likelihood (ML) estimator, restricted maximum likelihood (REML) estimator, and Sidik-Jonkman (SJ) estimator. HE, DL, and SJ estimator are originally proposed in random-effects meta-analysis but can be easily extended to the meta-regression context as shown in Section 2. Additionally, we consider an iterated (IT) version of the Sidik-Jonkman estimator and the Mandel-Paule (MP) estimator in the meta-regression context and, interestingly, we can prove that these two estimators are identical.

The outline of this paper is as follows: Section 2 contains the description of the general random-effects meta-regression model and the residual heterogeneity estimators. In Section 3, we present results of a numerical study. We discuss properties (bias and mean squared error (MSE)) of the estimators of residual heterogeneity and of the regression parameters. Further, properties of the confidence intervals of the regression parameters are examined. Finally, some concluding remarks are given.

Throughout this paper, we will use the following matrix notation: Let A be a matrix then A^T is the transposed matrix of A. Let A be a regular matrix then A^{-1} denotes the inverse of A. For a quadratic matrix A, let $|A|$ denote the determinant of A and $\mathrm{tr}(A)$ the trace of A. 1_k is the k-dimensional vector consisting of ones.

2. Estimators of Residual Heterogeneity

Let us consider k independent studies. The general random-effects meta-regression model can be written as:

$$\widehat{\theta}_i \sim \mathcal{N}\left(x_i^T \beta, \tau^2 + \widehat{\sigma}_i^2(\widehat{\theta}_i)\right), \quad i = 1, \ldots, k, \quad (1)$$

where $\widehat{\theta}_i$ is the estimator of the study-specific parameter of interest, $x_i \in \mathbb{R}^p$, $p < k$, the vector of known covariates, where the first component is one, $\beta \in \mathbb{R}^p$ the vector of unknown regression parameters, $\tau^2 \geqslant 0$ the residual heterogeneity parameter, and $\widehat{\sigma}_i^2(\widehat{\theta}_i)$ the estimated variance of $\widehat{\theta}_i$ in the ith study treated as fixed. In the following, we simply write $\widehat{\sigma}_i^2$ for the estimated within-study variability. If no covariates are used, that is, $x_i = 1$, we have the usual random-effects meta-analysis model with $\beta \in \mathbb{R}$ the parameter of interest.

In matrix notation, model (1) can be written as

$$\widehat{\theta} \sim \mathcal{N}(X\beta, \Lambda), \quad \Lambda = \tau^2 I_k + \widehat{\Delta}, \quad (2)$$

where $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)^T \in \mathbb{R}^k$ is the observable vector of study-specific estimators, $X \in \mathbb{R}^{k \times p}$ is the known matrix of covariates, whose ith row contains the covariate values of the ith study, and has full column rank $rk(X) = p$, $p < k - 1$, $\beta \in \mathbb{R}^p$ is the unknown vector of regression parameters, $\tau^2 \geqslant 0$ the residual heterogeneity parameter, I_k the $(k \times k)$-identity matrix, and $\widehat{\Delta} = \mathrm{diag}\{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2\}$ the $(k \times k)$-diagonal matrix with the estimated within-study variances as entries on the diagonal. Recall that for $X = 1_k$, we have the usual random-effects meta-analysis model in matrix notation.

In the following, we use either the notation of (1) or (2) based on convenience. We consider estimators of τ^2 based either on residual sum of squares or on likelihood approaches.

Given an estimator of the residual heterogeneity parameter τ^2 as described below, say $\widehat{\tau}^2$, we obtain the estimated covariance matrix $W = \widehat{\tau}^2 I_k + \widehat{\Delta}$. The weighted least squares estimator of β is then $\widehat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} \widehat{\theta}$ with $\widehat{\mathrm{Cov}}(\widehat{\beta}) =$

$(X^T W^{-1} X)^{-1}$. An approximate $(1-\alpha)$-confidence interval on a single regression parameter β_j, $j = 1, \ldots, p$, is given by

$$\widehat{\beta}_j \pm \sqrt{v_{jj}} z_{1-\alpha/2}, \tag{3}$$

where $\widehat{\beta}_j$ is the jth component of $\widehat{\beta}$, v_{jj} is the jth diagonal element of $\widehat{\mathrm{Cov}}(\widehat{\beta})$, and z_γ is the γ-quantile of the standard normal distribution, see e. g. (Ref. 2)

2.1. Hedges-type estimator

Consider the vector of ordinary least squares residuals in model (2)

$$(I_k - X(X^T X)^{-1} X^T)\widehat{\theta} = P_1 \widehat{\theta}$$

with $P_1 = I_k - X(X^T X)^{-1} X^T$. Note that P_1 is symmetric and idempotent. The residual sum of squares is given by $\widehat{\theta}^T P_1 \widehat{\theta}$. Under model (2) we have $\mathrm{E}(\widehat{\theta}^T P_1 \widehat{\theta}) = \mathrm{tr}(P_1 \mathrm{Cov}(\widehat{\theta}))$, since $P_1 X = 0$. Now $\mathrm{tr}(P_1 \mathrm{Cov}(\widehat{\theta})) = \tau^2 \mathrm{tr}(P_1) + \mathrm{tr}(P_1 \widehat{\Delta}) = \tau^2(k-p) + \mathrm{tr}(P_1 \widehat{\Delta})$.
The method of moments estimating equation is then $\widehat{\theta}^T P_1 \widehat{\theta} = \tau^2(k-p) + \mathrm{tr}(P_1 \widehat{\Delta})$ leading to

$$\widehat{\tau}_{HED}^2 = \frac{1}{k-p}\left(\widehat{\theta}^T P_1 \widehat{\theta} - \mathrm{tr}(P_1 \widehat{\Delta})\right). \tag{4}$$

Note that $\widehat{\tau}_{HED}^2$ can yield negative estimates of the nonnegative residual heterogeneity parameter and is usually truncated at 0. If $X = 1_k$, then $\widehat{\tau}_{HED}^2$ reduces to the estimator of heterogeneity introduced in the random-effects meta-analysis model[4].

2.2. DerSimonian-Laird-type estimator

This part is patterned from (Ref. 5). Consider the vector of weighted least squares residuals under $H_0 : \tau^2 = 0$ in model (2):

$$\left(I_k - X\left(X^T \widehat{\Delta}^{-1} X\right)^{-1} X^T \widehat{\Delta}^{-1}\right)\widehat{\theta} = P_2 \widehat{\theta}$$

with $P_2 = I_k - X\left(X^T \widehat{\Delta}^{-1} X\right)^{-1} X^T \widehat{\Delta}^{-1}$.

Note that P_2 is idempotent. The weighted residual sum of squares is then $\widehat{\theta}^T P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\theta}$.

Under model (2), we obtain with analogue calculations as above

$$\mathrm{E}(\widehat{\theta}^T P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\theta})$$
$$= \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2 \mathrm{Cov}(y))$$
$$= \tau^2 \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2) + \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\Delta}).$$

This leads to the method of moments estimating equation

$$\widehat{\theta}^T P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\theta} = \tau^2 \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2) + \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\Delta})$$

which reveals the estimator

$$\begin{aligned} \widehat{\tau}_{DL}^2 &= \frac{\widehat{\theta}^T P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\theta} - \mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2 \Delta)}{\mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2)} \\ &= \frac{\widehat{\theta}^T P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\theta} - (k-p)}{\mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2)}. \end{aligned} \tag{5}$$

Note that $\widehat{\Delta}^{-1} P_2 \widehat{\Delta} = I_k - \widehat{\Delta}^{-1} X(X^T \widehat{\Delta}^{-1} X)^{-1} X^T$ such that

$$P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\Delta} = I_k - \widehat{\Delta}^{-1} X(X^T \widehat{\Delta}^{-1} X)^{-1} X^T$$

and

$$\mathrm{tr}(P_2^T \widehat{\Delta}^{-1} P_2 \widehat{\Delta}) = k - p.$$

Like the Hedges-type estimator, the DerSimonian-Laird-type estimator can yield negative estimates of the nonnegative residual heterogeneity parameter and is, thus, truncated at zero in practice. If $X = 1_k$, then $\widehat{\tau}_{DL}^2$ reduces to the estimator of heterogeneity introduced in the random-effects meta-analysis model[6].

2.3. Mandel-Paule-type estimator

Let be $\Lambda = \tau^2 I_k + \widehat{\Delta}$ the covariance matrix in model (2) and consider the generalized least squares vector of residuals

$$(I_k - X(X^T \Lambda^{-1} X)^{-1} X^T \Lambda^{-1})\widehat{\theta} = P_3 \widehat{\theta}$$

with $P_3 = I_k - X(X^T \Lambda^{-1} X)^{-1} X^T \Lambda^{-1}$. Note that P_3 is idempotent. The weighted residual sum of squares is $\widehat{\theta}^T P_3^T \Lambda^{-1} P_3 \widehat{\theta}$. Under normality, it holds $\widehat{\theta}^T P_3^T \Lambda^{-1} P_3 \widehat{\theta} \sim \chi_{k-p}^2$. Recall $\mathrm{E}(\chi_{k-p}^2) = k-p$. Note that the only unknown parameter in Λ is τ^2. Consider $Q(\tau^2) = \widehat{\theta}^T P_3^T \Lambda^{-1} P_3 \widehat{\theta}$ as a function of τ^2.

This function is monotone decreasing in τ^2 as shown in Appendix A.

The Mandel-Paule-type estimator $\widehat{\tau}^2_{MP}$ is then defined as the unique solution of the equation

$$Q(\widehat{\tau}^2_{MP}) = k - p, \qquad (6)$$

if the solution exists, or $\widehat{\tau}^2_{MP} = 0$ if $Q(0) < k - p$.

If $X = 1_k$, then $p = 1$, and Eq. (6) reduces to the estimating equation originally proposed in the random-effects meta-analysis model[7]. Appendix B contains R code showing that the Mandel-Paule-type estimator can be easily calculated within the R package `metafor`.

2.4. Sidik-Jonkman-type estimator

Let us rewrite model (2) as

$$\widehat{\theta} \sim \mathcal{N}\left(X\beta, \tau^2 V\right) \text{ with } V = \text{diag}\left(1 + \frac{\widehat{\sigma}_i^2}{\tau^2}\right)_{i=1,\dots,k}. \qquad (7)$$

Then the vector of weighted least squares residuals is given by

$$(I_k - X(X^T V^{-1} X)^{-1} X^T V^{-1})\widehat{\theta} = P_4 \widehat{\theta}$$

with $P_4 = I_k - X(X^T V^{-1} X)^{-1} X^T V^{-1}$. Note that P_4 is idempotent. The weighted residual sum of squares is then

$$\begin{aligned} & \widehat{\theta}^T P_4^T V^{-1} P_4 \widehat{\theta} \\ = \ & \widehat{\theta}^T \left(V^{-1} - V^{-1} X(X^T V^{-1} X)^{-1} X^T V^{-1}\right) \widehat{\theta}. \end{aligned}$$

Then it holds $\mathrm{E}(\widehat{\theta}^T P_4^T V^{-1} P_4 \widehat{\theta}) = \tau^2 (k - p)$. This leads to an unbiased estimator of τ^2, namely

$$\widetilde{\tau}^2 = \frac{1}{k-p} \widehat{\theta}^T P_4^T V^{-1} P_4 \widehat{\theta}, \qquad (8)$$

which, in contrast to the above discussed estimators, always yields positive values (with probability 1) for the residual heterogeneity parameter. But $\widetilde{\tau}^2$ depends through V on the unknown τ^2. In the meta-analysis context, (Ref. 8)

$$\tau_0^2 = \frac{1}{k} \sum_{i=1}^{k} \left(\widehat{\theta}_i - \frac{1}{k} \sum_{j=1}^{k} \widehat{\theta}_j\right)^2 \qquad (9)$$

as the crude estimate for between-study variability. In meta-regression context, (Ref. 9) use the same crude estimator of τ^2 and plug this estimate in $\widetilde{\tau}^2$ though the meaning of τ^2 in meta-regression is different from the meaning of τ^2 in meta-analysis. In the following, we will denote the (one-step) Sidik-Jonkman-type estimator by $\widehat{\tau}^2_{SJ}$ when we replace τ^2 by τ_0^2 in $\widetilde{\tau}^2$. Note that, if the crude estimator τ_0^2 is equal to zero, then $\widehat{\tau}^2_{SJ} = 0$.

2.5. Iterated Sidik-Jonkman-type estimator

At least in meta-regression, it is appealing to iterate the Sidik-Jonkman-type estimator that is, we consider in the jth iteration step

$$\widehat{\tau}^2_j = \frac{1}{k-p} \widehat{\theta}^T \widehat{P}_{4,j-1}^T \widehat{V}_{j-1}^{-1} \widehat{P}_{4,j-1} \widehat{\theta}, \qquad (10)$$

starting with τ_0^2 from Eq. (9) on the right-hand side of Eq. (10). $\widehat{P}_{4,j-1}$ and $\widehat{V}_{j-1}^{-1}$ are the corresponding matrices with τ^2 replaced by $\widehat{\tau}^2_{j-1}$. We continue the iteration until convergence and denote the final estimator by $\widehat{\tau}^2_{IT}$. Note that $\widehat{\tau}^2_{IT}$ is set to zero if $\tau_0^2 = 0$.

Lemma 1. *The Mandel-Paule-type estimator $\widehat{\tau}^2_{MP}$ and the iterated Sidik-Jonkman-type estimator $\widehat{\tau}^2_{IT}$ are identical.*

Proof. Based on the unbiased estimator $\widetilde{\tau}^2$, see Eq. (8), we obtain the estimating equation for the iterated Sidik-Jonkman-type estimator as

$$\tau^2 = \frac{1}{k-p} \widehat{\theta}^T P_4^T V^{-1} P_4 \widehat{\theta}.$$

Note $P_4^T V^{-1} P_4 = V^{-1} - V^{-1} X(X^T V^{-1} X)^- X^T V^{-1}$. Using Theorem 1 of Appendix A, the above estimating equation can be rewritten as

$$\tau^2 = \frac{1}{k-p} \widehat{\theta}^T K (K^T V K)^{-1} K^T \widehat{\theta}$$

with $K^T X = 0$ and K^T of maximum row rank. Recall that $\tau^2 V = \Lambda$ and rearranging the last estimating equation to

$$\begin{aligned} k - p &= \frac{1}{\tau^2} \widehat{\theta}^T K (K^T V K)^{-1} K^T \widehat{\theta} \\ &= \widehat{\theta}^T K (K^T \tau^2 V K)^{-1} K^T \widehat{\theta} \\ &= \widehat{\theta}^T K (K^T \Lambda K)^{-1} K^T \widehat{\theta} \end{aligned}$$

yields the estimating equation for the Mandel-Paule-type estimator. This completes the proof. □

2.6. *Maximum Likelihood Estimation*

Under normality assumption of $\widehat{\theta}_i$ consider the conditional log-likelihood function

$$\ell(\beta, \tau^2 | \widehat{\theta}_1, \ldots, \widehat{\theta}_k, \widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2)$$
$$= \text{const} - \frac{1}{2} \sum_{i=1}^{k} \ln(\tau^2 + \widehat{\sigma}_i^2) - \frac{1}{2} \sum_{i=1}^{k} \frac{(\widehat{\theta}_i - x_i^T \beta)^2}{\tau^2 + \widehat{\sigma}_i^2}.$$

The partial derivatives of ℓ with respect to β and τ^2 are, defining $w_i = (\tau^2 + \widehat{\sigma}_i^2)^{-1}$, $i = 1, \ldots, k$,

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial}{\partial \beta} \left(-\frac{1}{2} \sum_{i=1}^{k} w_i \left(\widehat{\theta}_i - x_i^T \beta \right)^2 \right)$$
$$= \sum_{i=1}^{k} (w_i \widehat{\theta}_i) x_i - \sum_{i=1}^{k} [w_i (x_i^T \beta)] x_i$$

and

$$\frac{\partial \ell}{\partial \tau^2} = -\frac{1}{2} \sum_{i=1}^{k} \frac{1}{\tau^2 + \widehat{\sigma}_i^2} + \frac{1}{2} \sum_{i=1}^{k} \frac{(\widehat{\theta}_i - x_i^T \beta)^2}{(\tau^2 + \widehat{\sigma}_i^2)^2}.$$

Setting both partial derivatives equal to zero yield the estimation equations

$$\sum_{i=1}^{k} w_i (x_i^T \beta) x_i = \sum_{i=1}^{k} (w_i \widehat{\theta}_i y_i) x_i$$

and

$$\tau^2 = \sum_{i=1}^{k} w_i^2 \left[(y_i - x_i^T \beta)^2 - \widehat{\sigma}_i^2 \right],$$

which have to be solved iteratively under the constraint of nonnegative τ^2.

2.7. *Restricted Maximum Likelihood Estimation*

The maximum likelihood estimator of τ^2 tends to underestimate the residual heterogeneity in finite samples by failing to account for the fact that the regressions parameters are also estimated from the data. The restricted maximum likelihood estimator compensates for this underestimation. Following

(Ref. 10)) and using the model in matrix notation from Eq. (2), the log-likelihood function to be maximized is then

$$\ell_R(\tau^2 | \widehat{\theta}_1, \ldots, \widehat{\theta}_k, \widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2)$$
$$= -\frac{1}{2} \ln |\Lambda| - \frac{1}{2} |X^T \Lambda^{-1} X|$$
$$\quad -\frac{1}{2} (\widehat{\theta} - X \widetilde{\beta})^T \Lambda^{-1} (\widehat{\theta} - X \widetilde{\beta}),$$

leaving out the additive constant and $\widetilde{\beta}$ is the maximum likelihood solution of β given fixed τ^2 (Ref. 11). This log-likelihood function has to be maximized numerically. In the R package `metafor`, Fisher scoring algorithm is implemented for maximizing this function. This algorithm is known to be robust to poor starting values and usually converges quickly.[11,12] By default, the starting value is set equal to the value of the Hedges-type estimator and the algorithm terminates when the change in the estimated value of τ^2 is smaller than 10^{-5} from one iteration to the next. The Fisher scoring algorithm makes use of step halving to guarantee a non-negative estimate of τ^2.

3. Numerical Study

In a numerical study, we investigate the properties of the residual heterogeneity estimators Section 2 and their influence on the regression parameter estimates. Motivated by real data examples[2,13], we do not generate the data from the model (1) or (2), respectively, but only analyze the data in this model. For generating the data, we use a binomial-normal hierarchical model. Let p_{Ti} and p_{Ci} be the success probabilities in the treatment (T) and control (C) group, respectively, in the ith study, $i = 1, \ldots, k$. We consider the relative risk as effect size of interest for comparing treatment versus control. Then $\theta_i = \ln(p_{Ti}) - \ln(p_{Ci})$ is the study-specific logarithmic (log) relative risk. Let us consider one covariate x_i, then θ_i should follow a normal distribution with mean $\alpha + \beta x_i$ and variance τ^2.

In the numerical study, we fix $\alpha = 0$ and $\beta = -0.37$; the value of the slope parameter is motivated by the real data example in (Ref. 13). The values of x_i are drawn from $\mathcal{N}(0, 0.3^2)$. We choose the

mean equal to 0 because the covariate is often centered around its mean to give the intercept α a meaningful interpretation in the model. We fix the success probabilities in the control groups to $p_{Ci} = 0.15$, $i = 1, \ldots, k$, which is the mean success probability of the control groups in (Ref. 13). The considered values of the residual heterogeneity τ^2 range from 0 to 0.5 by 0.05. The sample sizes in the studies are always balanced, that is, $n_{Ti} = n_{Ci} = n_i$, and drawn either from a uniform discrete distribution between 10 and 30 (*small* sample sizes) or from a uniform discrete distribution between 100 and 300 (*large* sample sizes). We report the results for $k = 10$ and $k = 30$ studies. In total, we consider 44 scenarios.

In summary, the binomial-hierarchial model for generating the data is given by

$$Y_{Ti} \sim Bin(n_{Ti}, p_{Ti})$$
$$Y_{Ci} \sim Bin(n_{Ci}, p_{Ci})$$
$$p_{Ci} = 0.15$$
$$p_{Ti} = \min\{p_{Ci}\exp(\alpha + \beta x_i + a_i); 0.9999\}$$
$$\theta_i = \ln(p_{Ti}) - \ln(p_{Ci}) = \alpha + \beta x_i + a_i$$
$$a_i \sim \mathcal{N}(0, \tau^2)$$

Note that, to guarantee $p_{Ti} \in (0,1)$, we truncate $p_{Ci}\exp(\alpha + \beta x_i + a_i)$ at 0.9999. After generating n_{Ti}, n_{Ci}, p_{Ti}, and p_{Ci}, we generate the number of successes Y_{Ti} and Y_{Ci} from binomial distributions. Then

$$\widehat{\theta}_i = \ln\left(\frac{Y_{Ti}}{n_{Ti}}\right) - \ln\left(\frac{Y_{Ci}}{n_{Ci}}\right)$$

with

$$\widehat{\sigma}_i^2 = \frac{1}{Y_{Ti}} - \frac{1}{n_{Ti}} + \frac{1}{Y_{Ci}} - \frac{1}{n_{Ci}}.$$

The numerical study is carried out using the statistic software R (Ref. 14). Each estimated bias and mean squared error of the estimators is based on 2,000 simulation runs. Each estimated confidence coefficient is based on 10,000 simulation runs.

In the following sections, we report the fundamental results of the numerical study. We use the abbreviations HE, DL, ML, REML, SJ, and MP standing for Hedges-type estimator, DerSimonian-Laird-type estimator, maximum likelihood estimator, restricted maximum likelihood estimator, Sidik-Jonkman-type estimator, and Mandel-Paule-type estimator. Also we will use these abbreviations when the corresponding residual heterogeneity estimator is used in estimating a regression parameter or calculating a confidence interval for a regression parameter.

3.1. Properties of the Estimators of Residual Heterogeneity

In Figure 1, the biases of the residual heterogeneity estimators are displayed for $k = 10$ and $k = 30$ studies with *small* and *large* sample sizes, respectively.

For $k = 10$ studies and *small* sample sizes, SJ always has a positive bias, which is decreasing for increasing τ^2 and lies roughly between 0.35 and 0.2. All the other estimators have positive bias for small τ^2 and negative bias for larger τ^2. The biases of these estimators as functions of the residual heterogeneity parameter are all decreasing for increasing τ^2. HE is positively biased up to roughly $\tau^2 = 0.2$ and then negatively biased. MP, DL, and REML are positively biased close to zero and then all are negatively biased. Within this group of estimators, MP performs best. ML has the largest negative bias.

For $k = 30$ studies and *small* sample sizes, we obtain practically the same results as for $k = 10$ studies. ML has again the largest negative bias but the amount of bias is less than for $k = 10$ studies. HE is now positively biased only close to $\tau^2 = 0$ and then negatively biased. The amount of bias of HE is less than the biases of MP, DL, and REML. With respect to bias, given small sample sizes, HE seems to perform best followed by MP, DL, and REML. ML cannot be recommended because of its downward bias. SJ is not suitable for small sample sizes.

By increasing the sample sizes, the amount of biases decreases as expected. All the biases of the estimators, except ML, tend to be close to zero. For $k = 10$ studies, REML seems to be perform best followed by MP and DL, where DL is negatively biased for large τ^2 and MP positively biased for larger τ^2. SJ and HE are always positively biased, where the amount of bias is larger with HE than SJ when the residual heterogeneity is large. For $k = 30$ studies, MP seems to be perform best followed by REML. DL and ML are negatively biased, HE and SJ again positively biased.

In Figure 2, mean squared errors (MSE) of the

k = 10 and small sample sizes

k = 30 and small sample sizes

k = 10 and large sample sizes

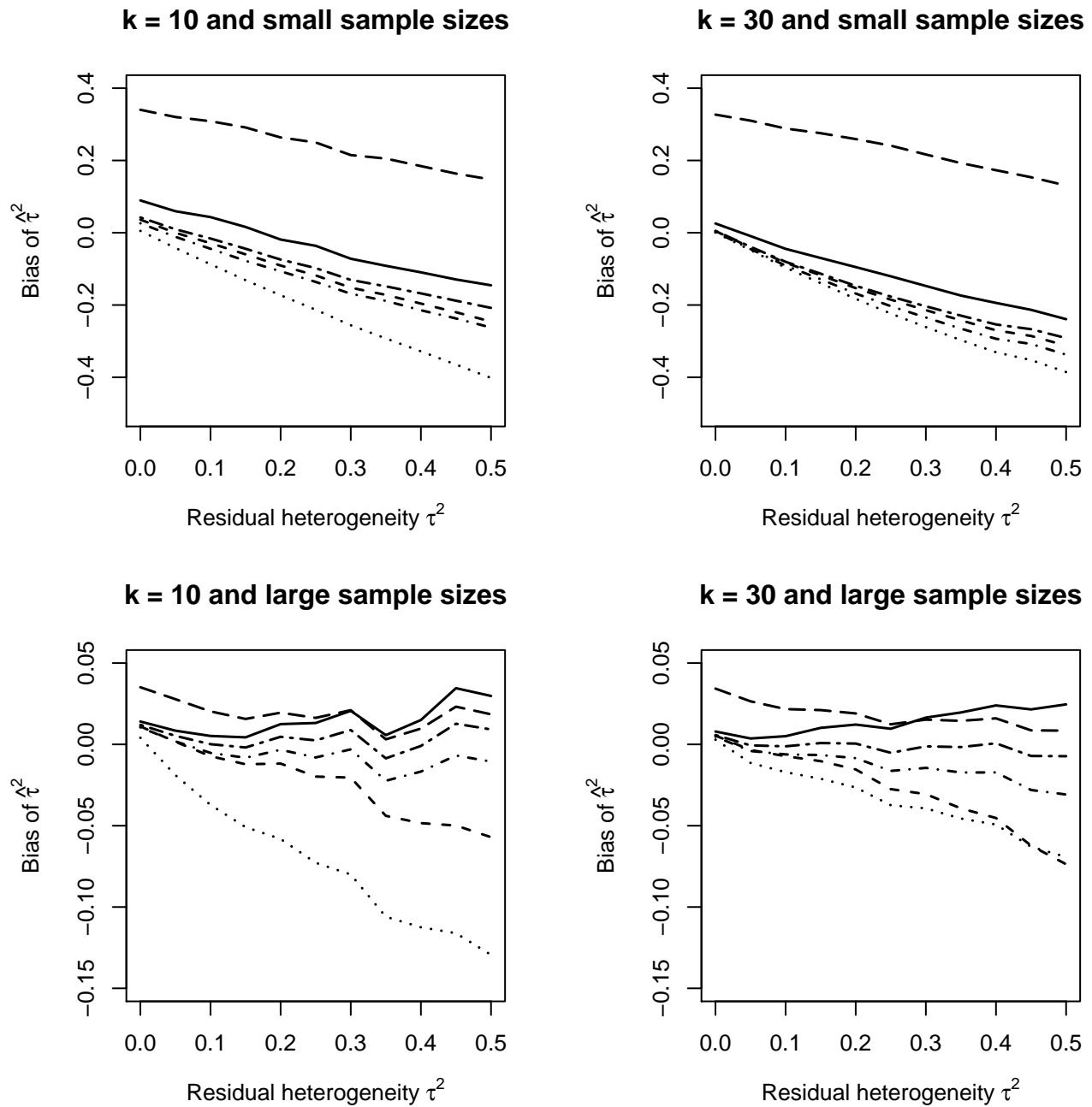**k = 30 and large sample sizes**

Figure 1: Biases of the residual heterogeneity estimators HE (solid line), DL (dashed), ML (dotted), REML (dotdash), SJ (longdash), and MP (twodash) for $k = 10$ and $k = 30$ studies with *small* and *large* sample sizes
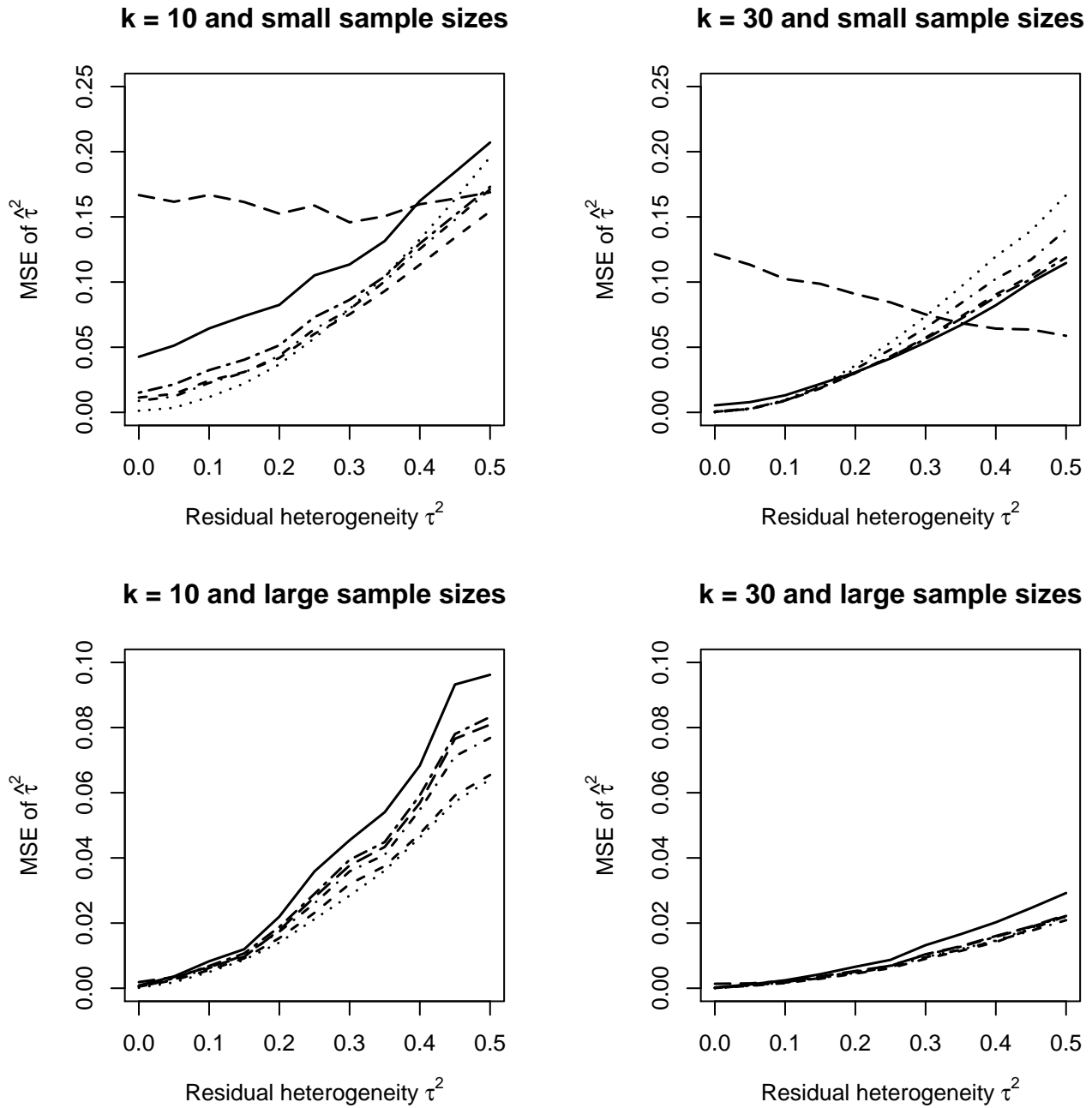
Figure 2: Mean squared errors (MSE) of the residual heterogeneity estimators HE (solid line), DL (dashed), ML (dotted), REML (dotdash), SJ (longdash), and MP (twodash) for $k = 10$ and $k = 30$ studies with *small* and *large* sample sizes

residual heterogeneity estimators are displayed for $k = 10$ and $k = 30$ studies with *small* and *large* sample sizes, respectively.

For small sample sizes, SJ shows a different behavior than the other five estimators. Given $k = 10$ studies, the MSE of SJ is roughly constant over the range of considered heterogeneity values. Up to $\tau^2 = 0.35$, SJ has the largest MSE of all estimators. Looking at the other five estimators, the MSEs are all increasing with increasing residual heterogeneity τ^2. HE clearly has the largest MSE in this group and, for large values of τ^2, even a larger MSE than SJ; the other four estimators nearly have identical MSEs. For $k = 30$ studies, the MSE's of all estimators are smaller than for $k = 10$ studies. SJ has now decreasing MSE with increasing τ^2, and, up to $\tau^2 = 0.3$, SJ has the largest MSE of all estimators. The other five estimators again have increasing MSEs with increasing τ^2. In most cases, ML has the largest MSE in this group, and for $\tau^2 > 0.2$, HE has the smallest MSE.

For large sample sizes, all the estimators possess increasing MSE with increasing residual heterogeneity. For $k = 10$ and $k = 30$, HE has the largest MSE, while all the other estimators have practically identical MSEs. For $k = 10$ and large values of τ^2, ML and DL have smaller MSEs than the other estimators. Increasing the number of studies from $k = 10$ to $k = 30$ leads to smaller MSEs of the estimators.

3.2. *Properties of the Estimators of Regression Parameters*

In the second step of our numerical study, we investigate the impact of the residual heterogeneity estimators on the properties of the regression parameter estimators. Interestingly, we do not observe substantial differences between the properties of the regression parameter estimates using the different residual heterogeneity estimators.

For *small* sample sizes, all the intercept estimators are positively biased. The biases are all increasing for increasing τ^2. SJ always yields the smallest bias, ML always the largest. The biases of the other four methods lie in between and are virtually identical. With increasing number of studies, the bi-

ases of the other four methods approach the bias of ML identifying SJ clearly as the less biased estimator. For large τ^2, the bias can be up to 0.2. For *large* sample sizes, the magnitude of the biases is, as expected, smaller and less than 0.04 even for large values of τ^2. All the methods produce practically identical biases. With respect to MSE, all methods yield nearly identical MSEs for $k = 10$ and $k = 30$ studies indicating that the variability of SJ may be larger than the variability of the other methods.

For the slope parameter, the biases are mostly positive and up to 0.05 given *small* sample sizes. Again SJ yields the smallest bias and all other methods produce nearly identical biases. For *large sample*, the biases of all methods are nearly identical and around zero. With respect to MSE, SJ has the largest MSE with *small* sample sizes indicating again that the variability of SJ is larger compared to the other estimators. The MSEs of the other methods are practically identical. With *large* samples sizes, all estimators yield virtually identical MSEs.

3.3. *Properties of the Confidence Intervals of Regression Parameters*

Using Eq. (3), we investigate the actual confidence coefficients of the confidence interval on the slope parameter β given a two-sided nominal level of 95%. In Table 1, the results are displayed for *small* sample sizes in $k = 10$ and $k = 30$ studies.

Looking at the results, we can first state that given a method and a fixed value of τ^2 there is practically no difference between the estimated confidence coefficients for $k = 10$ and $k = 30$ studies. Generally, the estimated confidence coefficients decrease for increasing τ^2. SJ produces always very conservative confidence intervals even for large values of τ^2. For small values of τ^2, all methods yield conservative confidence intervals. For larger values, first ML tends to produce liberal confidence intervals. In the group, HE, DL, REML, and MP, HE produces always the largest estimated confidence coefficient for a fixed value of τ^2 and shows no tendency to become liberal. MP has (nearly) always the second largest estimated confidence in this group and shows a tendency to become liberal only for large values of τ^2 with $k = 10$ studies. DL behave similar

to MP, and REML performs a little worse than MP and DL being a bit liberal for large values of τ^2.

Table 2 contains the results for *large* sample sizes. At a first glance, we see that the number of studies has an effect on the actual confidence coefficients. For $k = 10$ studies, the estimated confidence coefficient are obviously significantly less than for $k = 30$ studies. For $k = 10$ and residual heterogeneity present, all the methods produce very liberal confidence intervals except SJ for $\tau^2 = 0.05$. ML yields the worst results followed by DL especially for large values of τ^2. HE and MP, like for *small* sample sizes, produce similar results closely followed by REML. For $\tau^2 \geqslant 0.1$, where all the confidence intervals are liberal, SJ has always the largest estimated confidence coefficient. When no residual heterogeneity is present, all the methods yield conservative confidence intervals with SJ the most conservative one.

For $k = 30$ studies, all the methods again yield conservative confidence when no residual heterogeneity is present with SJ again the most conservative one. SJ remains a bit conservative for $\tau^2 \leqslant 0.2$ but for $\tau^2 > 0.2$ attains the nominal level. Second best method is HE with a slight tendency to be liberal for small values of τ^2. But even for small values of τ^2, HE is better than the remaining methods. MP closely follows HE as third best. ML is again the worst method when residual heterogeneity is present.

4. Concluding Remarks

Meta-regression has become a commonly used tool for investigating whether study characteristics may explain heterogeneity of results among studies in a meta-analysis. However, such explorations of heterogeneity are prone to misleading false-positive results. It is unclear how many covariates can be reliably investigated, and how this might depend on the number of studies, the extent of the heterogeneity and the relative weights awarded to the different studies.[15]

In this paper we have investigated six different residual heterogeneity estimators in random-effects meta-regression, four based on different residual sum of squares and two based on likelihood approaches. Estimators of residual heterogeneity are often used to quantify the reduction of heterogeneity due to the covariate(s). In our numerical study, we have seen that the estimators behave quite differently. Sidik-Jonkman-type and Hedges-type estimator tend to be positively biased even for large sample sizes. Consequently, quantifying the reduction of heterogeneity with these two estimators may lead to a different conclusion in comparison to, for instance, the DerSimonian-Laird-type estimator, which turned out to be negatively biased. For large samples sizes, the Mandel-Paule-type estimator seems to be nearly unbiased and is thus an attractive alternative to the already in the R package `metafor` implemented estimators.

With respect to confidence intervals on the regression parameters, we have seen that most of the confidence intervals are liberal for large sample sizes. To overcome this problem, the confidence intervals proposed in (Ref. 5) could be used.

Finally, let us recall that we have used a binomial-hierarchical model for generating the data but used model (2) for analyzing the data. Model (2) can be derived a the marginal model from a normal-normal hierarchical model. Our data generation process was motivated by the real data example and then we applied the standard analysis. Clearly, such data could also be analyzed using Bayesian approaches.[16]

References

1. L. V. Hedges and I. Olkin, "Statistical Methods for Meta-Analysis," Academic Press, Boston (1985).
2. J. Hartung, G. Knapp, and B. K. Sinha, "Statistical Meta-Analysis with Applications," Wiley, New York (2008).
3. W. Viechtbauer, "Conducting meta-analysis in R with the metafor package," *Journal of Statistical Software*, **36**, 1–48 (2010).
4. L. V. Hedges, "A random effects model for effect sizes," *Psychological Bulletin*, **93**, 388–395 (1983).

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

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

7. J. Mandel and R. C. Paule, "Interlaboratory evaluation of a material with unequal number of replicates," *Analytical Chemistry*, **42**, 1194–1197 (1970).

8. K. Sidik and J. N. Jonkman, "Simple heterogeneity variance estimation for meta-analysis", *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **54**, 367–384 (2005).

9. K. Sidik and J. N. Jonkman, "A note on variance estimation in random effects meta-regression," *Journal of Biopharmaceutical Statistics*, **15**, 823–838, (2005).

10. W. Viechtbauer, "Bias and efficiency of meta-analytic variance estimators in the random-effects model," *Journal of Educational and Behavioral Statistics*, **30**, 261–293 (2005).

11. D. A. Harville, "Maximum likelihood approaches to variance component estimation and to related problems," *Journal of the American Statistical Association*, **72**, 320–338 (1997).

12. R. I. Jennrich and P. F. Sampson, "Newton-Raphson and related algorithms for maximum likelihood variance component estimation," *Technometrics*, **18**, 11–17 (1976).

13. M. R. Law, N. J. Wald, and S. G. Thompson, "By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease?" *British Medical Journal*, **308**, 367–373 (1994).

14. R Development Core Team , "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/ (2012).

15. J. P. T. Higgins and S. G. Thompson, "Controlling the risk of spurious findings from meta-regression," *Statistics in Medicine*, **23**, 1663–1682 (2004).

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

17. G. C. Khatri, A note on a MANOVA model applied to problems in growth curves. *Ann. Inst. Stat. Math.*, **18**, 75–78 (1966).

18. J. R. Magnus and H. Neudecker, "Matrix Differential Calculus with Applications in Statistics and Econometrics," Wiley, Chichester (1988).

19. G. Knapp, B. J. Biggerstaff, and J. Hartung, "Assessing the amount of heterogeneity in random-effects meta-analysis," *Biometrical Journal*, **48**, 271–285 (2006).

20. W. Viechtbauer, "Confidence intervals for the amount of heterogeneity in meta-analysis," *Statistics in Medicine*, **26**, 37–52 (2007).

Appendix A

Uniqueness of Mandel-Paule-type estimator

Theorem 1. *(Khatri, 1966): If $K^T X = 0$, where K^T has maximum row rank, and V is positive definite then $K(K^T V K)^{-1} K^T = P$ for $P = V^{-1} - V^{-1} X (X^T V^{-1} X)^- X^T V^{-1}$.*

Using Theorem 1 we can write

$$
\begin{aligned}
P_3^T \Lambda^{-1} P_3 &= \Lambda^{-1} - \Lambda^{-1} X \left(X^T \Lambda^{-1} X\right)^{-1} X^T \Lambda^{-1} \\
&= K(K^T \Lambda K)^{-1} K^T
\end{aligned}
$$

with $K^T X = 0$ and K^T has maximum row rank. This leads to the quadratic from

$$
\begin{aligned}
Q(\tau^2) &= \widehat{\theta}^T P_3^T \Lambda^{-1} P_3 \widehat{\theta} = \widehat{\theta}^T K (K^T \Lambda K)^{-1} K^T \widehat{\theta} \\
&= \varepsilon^T (K^T (\tau^2 I_k + \widehat{\Delta}) K)^{-1} \varepsilon \\
&= \varepsilon^T (\tau^2 K^T K + K^T \widehat{\Delta} K)^{-1} \varepsilon
\end{aligned}
$$

with $\varepsilon = K^T \widehat{\theta}$. Following Magnus and Neudecker (1988, p. 151), we obtain

$$
\begin{aligned}
\frac{dQ(\tau^2)}{d\tau^2} &= \varepsilon^T \left(\frac{d}{d\tau^2} \left(\tau^2 K^T K + K^T \widehat{\Delta} K \right)^{-1} \right) \varepsilon \\
&= -\varepsilon^T \left(\tau^2 K^T K + K^T \widehat{\Delta} K \right)^{-1} K^T K \\
&\quad \times \left(\tau^2 K^T K + K^T \widehat{\Delta} K \right)^{-1} \varepsilon \\
&= -\eta^T K^T K \eta < 0, \\
&\qquad \text{since } K^T K \text{ is positive definite },
\end{aligned}
$$

with $\eta = \left(\tau^2 K^T K + K^T \widehat{\Delta} K \right)^{-1} \varepsilon \neq 0$. Consequently, $Q(\tau^2)$ is strictly monotone decreasing in the nonnegative residual heterogeneity parameter τ^2 and the Mandel-Paule-type estimator is unique.

Appendix B

R code for computing Mandel-Paule-type estimator

Knapp, Biggerstaff, and Hartung (2006) as well as Viechtbauer (2007) proposed an $(1 - \alpha)$-confidence interval on τ^2 defined by

$$
CI(\tau^2) = \left\{ \tau^2 | \chi^2_{k-p, \alpha/2} \leqslant Q(\tau^2) \leqslant \chi^2_{k-p, 1-\alpha/2} \right\}
$$

with $\chi^2_{\nu,\kappa}$ the κ-quantile of the χ^2-distribution with ν degrees of freedom. Since $Q(\tau^2)$ is monotone decreasing in τ^2, see Appendix A, the bounds of the confidence interval can be determined by solving the following two equations:

$$\text{Lower bound: } Q(\tau^2_{LB}) = \chi^2_{k-p,1-\alpha/2} \quad ;$$

$$\text{Upper bound: } Q(\tau^2_{UB}) = \chi^2_{k-p,\alpha/2}.$$

The computation of this interval, also called Q-profiling confidence interval, is implemented in the R function *confint.rma.uni* of the R package `metafor`.

For computing the Mandel-Paule-type estimator, one has to solve the equation $Q(\tau^2_{MP}) = k - p$, where $k - p$ is the mean of the χ^2-distribution with $k - p$ degrees of freedom. Since the mean of χ^2_{k-p} is greater than the median of χ^2_{k-p}, we first compute which quantile of χ^2_{k-p} is equal to the mean, say $\chi^2_{k-p;q}$, $q \geqslant 0.5$. The level of the confidence interval is the $1 - \alpha = 2 \times (q - 0.5)$ and the Mandel-Paule estimator is then the lower bound of the Q-profiling confidence interval at level $(1 - \alpha)$.

The R code for computing the Mandel-Paule estimator is given below.

```
MP      <- function(rma.uni.object){
# no. of studies
k       <- rma.uni.object$k
# no. of columns of X
p       <- rma.uni.object$p
# level of the confidence interval in %
level  <- 100*(2*(pchisq(k-p,k-p)-.5))
# upper bound of the Q-profiling c.i.
tau.MP<-confint.rma.uni(rma.uni.object,
                    level=level)[1,2]
return(tau.MP) }
```

Prerequisite: An *rma.uni.object* must exist, that is, one has already conducted a random-effects meta-analysis or meta-regression. Note that always the Q-profiling confidence interval is calculated irrespective of the (residual) heterogeneity estimator used for carrying out the meta-analysis or meta-regression.

Table 1: Actual coverage probabilities (in %) for the confidence intervals on β using different heterogeneity estimators with different values of τ^2 and sample sizes between 10 and 30 from $k = 10$ studies at nominal level of 95%

					Number of studies $k = 10$						
Method	$\tau^2 = 0$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
HE	98.17	98.15	97.72	97.26	96.82	96.66	96.21	95.66	94.95	94.96	94.72
DL	97.98	97.99	97.45	97.01	96.39	96.37	95.82	95.40	94.73	94.68	94.46
ML	97.90	97.79	97.24	96.66	96.05	95.78	94.98	94.33	93.34	93.35	92.87
REML	97.95	97.90	97.28	96.91	96.30	96.14	95.56	95.08	94.30	94.42	94.04
SJ	99.32	99.42	99.23	99.09	98.87	98.78	98.63	98.60	98.30	98.11	98.13
MP	98.01	97.98	97.48	97.06	96.41	96.42	95.85	95.43	94.79	94.64	94.57
					Number of studies $k = 30$						
Method	$\tau^2 = 0$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
HE	98.39	97.92	97.70	97.23	96.97	96.51	96.10	95.39	95.57	95.14	95.39
DL	98.24	97.73	97.37	96.99	96.57	95.93	95.44	94.80	95.10	94.69	95.01
ML	98.22	97.66	97.29	96.72	96.20	95.35	94.65	94.06	93.97	93.70	93.84
REML	98.22	97.68	97.26	96.79	96.35	95.56	94.97	94.45	94.67	94.32	94.51
SJ	99.62	99.44	99.43	99.25	99.07	99.07	98.84	98.76	98.73	98.46	98.46
MP	98.24	97.73	97.37	97.03	96.59	95.98	95.56	94.93	95.15	94.79	95.04

Table 2: Actual coverage probabilities (in %) for the confidence intervals on β using different heterogeneity estimators with different values of τ^2 and sample sizes between 100 and 300 from $k = 10$ studies at nominal level of 95%

					Number of studies $k = 10$						
Method	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
HE	96.16	92.93	91.90	91.95	91.78	91.87	92.14	92.14	91.82	92.39	91.99
DL	96.03	92.71	91.59	91.65	91.35	91.02	91.26	91.06	90.70	91.20	90.31
ML	95.52	90.48	88.63	88.17	88.22	87.97	88.18	88.27	88.00	88.47	87.84
REML	95.99	92.60	91.60	91.66	91.44	91.30	91.68	91.61	91.27	91.79	91.26
SJ	98.26	96.19	94.46	94.14	93.67	93.31	93.40	93.06	92.66	93.14	92.60
MP	96.07	92.79	91.75	91.90	91.81	91.64	92.06	91.94	91.59	92.19	91.75
					Number of studies $k = 30$						
Method	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
HE	96.55	94.39	94.12	94.52	94.85	94.75	94.81	94.66	94.81	95.09	95.16
DL	96.23	94.17	93.45	93.56	93.82	93.64	93.59	93.63	93.17	93.28	93.35
ML	95.91	93.05	92.58	92.77	93.14	93.07	93.13	93.24	93.05	93.23	93.34
REML	96.15	94.05	93.47	93.79	94.09	93.96	94.07	94.04	93.97	94.20	94.28
SJ	98.66	96.91	95.66	95.44	95.58	95.18	95.10	94.86	94.94	95.08	95.20
MP	96.28	94.33	93.81	94.20	94.49	94.40	94.39	94.30	94.34	94.61	94.70