Median bias reduction in random-effects meta-analysis and meta-regression

Sophia Kyriakou,[1] Ioannis Kosmidis[2,3] and Nicola Sartori[4]

Abstract

The reduction of the mean or median bias of the maximum likelihood estimator in regular parametric models can be achieved through the additive adjustment of the score equations. In this paper, we derive the adjusted score equations for median bias reduction in random-effects meta-analysis and meta-regression models and derive efficient estimation algorithms. The median bias-reducing adjusted score functions are found to be the derivatives of a penalised likelihood. The penalised likelihood is used to form a penalised likelihood ratio statistic which has known limiting distribution and can be used for carrying out hypothesis tests or for constructing confidence intervals for either the fixed-effect parameters or the variance component. Simulation studies and real data applications are used to assess the performance of estimation and inference based on the median bias-reducing penalised likelihood and compare it to recently proposed alternatives. The results provide evidence on the effectiveness of median bias reduction in improving estimation and likelihood-based inference.

Keywords

Adjusted score equations, heterogeneity, mean bias reduction, penalised likelihood, random effects

1 Introduction

Meta-analysis is a core tool for synthesising the results from independent studies investigating a common effect of interest. One of the main challenges when combining results from multiple studies is the variability or heterogeneity in the design and the methods employed in each study. Accounting for and quantifying that heterogeneity is critical when drawing inferences about the common effect. In this direction, DerSimonian and Laird[1] introduced the random-effects meta-analysis model, which expresses the heterogeneity between studies in terms of a variance component that can be estimated through standard estimation techniques.

Nevertheless, there is ample evidence that frequentist inference based on random-effects meta-analysis can be problematic in the usual meta-analytic scenario where the number of studies is small or moderate. Specifically, the estimation of the heterogeneity parameter can be highly imprecise, which in turn results in misleading conclusions.[2–4] Examples of recently proposed methods that attempt to improve inference are the resampling[5] and double resampling approaches,[6] and the mean bias-reducing penalised likelihood (BRPL) approach in Kosmidis et al.[3] Specifically, Kosmidis et al.[3] show that maximisation of the BRPL results in an estimator of the heterogeneity parameter that has notably smaller bias than maximum likelihood (ML) with small loss in efficiency, and illustrate that BRPL-based inference outperforms its competitors in terms of inferential performance.

Kenne Pagui et al.[7] show that under suitable conditions, third-order median unbiased estimators can be obtained by the solution of a suitably adjusted score equation. The components of such median bias-reduced estimators have, to third-order, the same probability of over- and under-estimating the true parameter. A key

[1]Department of Statistical Science, University College London, London, UK
[2]Department of Statistics, University of Warwick, Coventry, UK
[3]The Alan Turing Institute, London, UK
[4]Department of Statistical Sciences, University of Padova, Padova, Italy

Corresponding author:
Sophia Kyriakou, Department of Statistical Science, University College, 1-19 Torrington Place, London WC1E 7HB, UK.
Email: sophia.kyriakou.14@ucl.ac.uk

property of these estimators, not shared with the mean bias-reduced ones, is that any monotone component-wise transformation of the estimators results automatically in median bias-reduced estimators of the transformed parameters.[7] Such equivariance property can be useful in the context of random-effects meta-analysis where the Fisher information and, hence, the asymptotic variances of various likelihood-based estimators depend only on the heterogeneity parameter.

In this paper, we derive the median bias-reducing adjusted score functions for random-effects meta-analysis and meta-regression. The adjusted score functions are found to correspond to a median BRPL, whose logarithm differs from the logarithm of the mean BRPL in Kosmidis et al.[3] by a simple additive term that depends on the heterogeneity parameter. Since the adjustments to the score function for mean and median bias reduction are both of order $O(1)$, the same arguments as in Kosmidis et al.[3] are used to obtain a median BRPL ratio statistic with known asymptotic null distribution that can be used for carrying out hypothesis tests and constructing confidence regions or intervals for either the fixed-effect or the heterogeneity parameter. Simulation studies and real data applications are used to assess the performance of estimation and inference based on the median BRPL, and compare it to recently proposed alternatives, including the mean BRPL. The results provide evidence on the effectiveness of median bias reduction in improving estimation and likelihood-based inference.

The rest of the paper is organised as follows. Section 2 considers the cocoa intake dataset[8] and a random-effects meta-analysis model as a motivational example that demonstrates how conclusions in frequentist inference may vary using different methods when the number of studies is small. Section 3 defines the random-effects meta-analysis and meta-regression model and establishes notation. Section 4 gives a formal statement of the proposed median bias-reducing adjusted score function and penalised likelihood for these models, gives the algorithm used for computing the median BRPL estimates, and briefly discusses the median BRPL ratio statistic used for inference. Section 5 revisits the example in Section 2 and gives some simulations comparing the median BRPL method to alternative approaches. Section 6 gives more extensive simulations under the random-effects meta-analysis model that evaluate the performance of median BRPL and compare it with that of ML and mean BRPL methods. An application to meat consumption data[9] and a random-effects meta-regression model is given in Section 7. Section 8 concludes with a brief discussion, and Appendix 1 contains some technical details on the derivation of the median bias-reducing adjusted score functions.

## 2 Cocoa intake and blood pressure reduction data

Consider the setting in Bellio and Guolo[10] who carry out a meta-analysis of five randomised controlled trials from Taubert et al.[8] on the efficacy of two weeks of cocoa consumption on lowering diastolic blood pressure. Figure 1 is a forest plot with the estimated mean difference in diastolic blood pressure before and after cocoa intake from each study, and the associated 95% Wald-type confidence intervals. Four out of the five studies reported a reduction of diastolic blood pressure from cocoa intake.

The random-effects meta-analysis model is used to synthesise the evidence from the five studies. In particular, let Y_i be the random variable representing the mean difference in the diastolic blood pressure after two weeks of cocoa intake in the ith study. We assume that $Y_1, \ldots, Y_5$ are independent random variables, where Y_i has a Normal distribution with mean the overall effect β and variance $\hat{\sigma}_i^2 + \psi$, with $\hat{\sigma}_i^2$ the estimated standard error of the effect from the ith study and ψ the heterogeneity parameter.

The forest plot in Figure 1 has been enriched with several nominally 95% confidence intervals for β using various alternative methods. As is apparent, the conclusions when testing the hypothesis $\beta = 0$ can vary depending on the method used. More specifically, the Wald test using the ML estimates, the DerSimonian and Laird method,[1] double resampling,[6] and the likelihood ratio (LR) test give evidence that there is a relationship between cocoa consumption and diastolic blood pressure, with p-values 0.005, 0.006, 0.016, 0.030, respectively. On the other hand, Knapp and Hartung's method,[11] the mean BRPL ratio,[3] the Bartlett-corrected LR,[12] and Skovgaard's test suggest that the evidence that cocoa consumption affects diastolic blood pressure is weaker, with p-values of 0.050, 0.053, 0.058, and 0.067, respectively.

3 Random-effects meta-regression model

Let y_i and $\hat{\sigma}_i^2$ denote the estimate of the effect from the ith study ($i = 1, \ldots, K$) and the associated within-study variance, respectively, and $x_i = (x_{i1}, \ldots, x_{ip})^T$ denote a p-vector of study-specific covariates that can be used to account for the heterogeneity across studies.
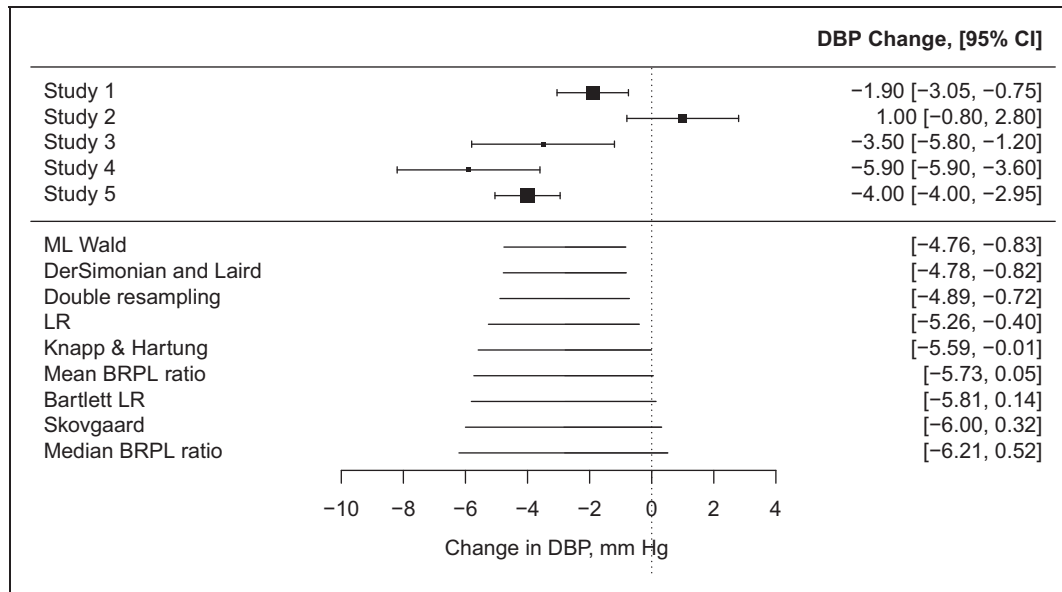
Figure 1. Forest plot of cocoa data.[8,10] The outcomes from the five studies are reported in terms of the diastolic blood pressure (DBP) difference after two weeks of cocoa consumption. A negative change in DBP indicates favorable hypotensive cocoa actions. Squares represent the mean effect estimate for each study; the size of the square reflects the weight that the corresponding study exerts in the meta-analysis calculated as the within-study's inverse variance. Horizontal line segments represent 95% Wald-type confidence intervals (CI) of the effect estimate of individual studies. In the bottom panel of the plot, horizontal line segments represent the corresponding 95% confidence interval as computed based on the Wald statistic using the ML estimates (ML Wald), the DerSimonian and Laird approach,[1] double resampling,[6] the LR statistic, the Knapp and Hartung[11] method, the mean BRPL ratio statistic,[3] the Bartlett-corrected LR statistic (Bartlett LR),[12] the Skovgaard's statistic, and the median BRPL ratio statistic. The confidence intervals *n* are ordered according to their length. The estimate of β has not been reported, as is commonly done in forest plots, because some of the methods considered (e.g. Skovgaard, Bartlett-corrected LR, and double resampling) are designed to produce directly *p*-values and/or confidence intervals and do not directly correspond to an estimation method.

The within-study variances $\hat{\sigma}_i^2$ are usually assumed to be estimated well enough to be considered as known and equal to the values reported in each study. Then the observations $y_1, \ldots, y_K$ are assumed to be realisations of the random variables $Y_1, \ldots, Y_K$, which are independent conditionally on independent random effects $U_1, \ldots, U_K$. The conditional distribution of Y_i given $U_i = u_i$ is $N(u_i + x_i^T \beta, \hat{\sigma}_i^2)$, where β is an unknown *p*-dimensional vector of fixed effects. The random effects $U_1, \ldots, U_K$ are typically assumed to be independent with U_i having a $N(0, \psi)$ distribution, where ψ is a parameter that attempts to capture the unexplained between-study heterogeneity. In matrix notation, the random-effects meta-regression model has

$$Y = X\beta + U + \epsilon \tag{1}$$

where $Y = (Y_1, \ldots, Y_K)^T$, X is the $K \times p$ model matrix with x_i^T in its *i*th row, and $\epsilon = (\epsilon_1, \ldots, \epsilon_K)^T$ is a vector of independent errors each with a $N(0, \hat{\sigma}_i^2)$ distribution and independent of $U = (U_1, \ldots, U_K)^T$. Under this specification, the marginal distribution of Y is multivariate normal with mean $X\beta$ and variance–covariance matrix $\hat{\Sigma} + \psi I_K$, where I_K is the $K \times K$ identity matrix and $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2)$. The random-effects meta-analysis results as a special case of meta-regression, by setting X to be a column of ones.

The log-likelihood function for $\theta = (\beta^T, \psi)^T$ is $l(\theta) = \{\log|W(\psi)| - R(\beta)^T W(\psi) R(\beta)\}/2$, where $|W(\psi)|$ denotes the determinant of $W(\psi) = (\hat{\Sigma} + \psi I_K)^{-1}$ and $R(\beta) = y - X\beta$. The gradient of the log-likelihood (score function) is

$$s(\theta) = \begin{pmatrix} X^T W(\psi) R(\beta) \\ \frac{1}{2}\{R(\beta)^T W(\psi)^2 R(\beta) - \text{tr}[W(\psi)]\} \end{pmatrix} \tag{2}$$

and the ML estimator $\hat{\theta} = (\hat{\beta}^T, \hat{\psi})^T$ is obtained as the solution of $s(\theta) = 0_{p+1}$, where 0_p denotes a *p*-dimensional vector of zeros.

4 Median bias reduction

4.1 The method

A popular method for reducing the mean bias of ML estimates in regular statistical models is through the adjustment of the score equation.[13,14] Kenne Pagui et al.[7] propose an extension of the adjusted score equation approach which can be used to obtain median bias-reduced estimators. Specifically, under the model, the new estimator has a distribution with median closer to the "true" parameter value than the ML estimator. Kenne Pagui et al.[7] consider the median as a centering index for the score, and the adjusted score function for median bias reduction then results by subtracting from the score its approximate median, obtained using a Cornish-Fisher asymptotic expansion.

Let $j(\theta) = -\partial^2 l(\theta)/\partial\theta\partial\theta^{\mathrm{T}}$ be the observed information matrix (see Appendix 1 for its expression), and $i(\theta)$ be the expected information matrix

$$i(\theta) = E_\theta(j(\theta)) = \begin{pmatrix} X^{\mathrm{T}} W(\psi) X & 0_p \\ 0_p^{\mathrm{T}} & \frac{1}{2}\mathrm{tr}[W(\psi)^2] \end{pmatrix} \tag{3}$$

with tth column $i_t(\theta)$. Let also $i^t(\theta)$ and $i^{tt}(\theta)$ be the tth column and the tth diagonal element of $\{i(\theta)\}^{-1}$, with $t \in \{1, \ldots, p+1\}$. Kenne Pagui et al.[7] show that a median bias-reduced estimator $\hat{\theta}^\dagger$ can be obtained by solving an adjusted score equation of the form $s^\dagger(\theta) = s(\theta) + A^\dagger(\theta) = 0$, where the additive adjustment to $A^\dagger(\theta)$ is $O(1)$, in the sense that $A^\dagger(\theta)$ is bounded in absolute value by a fixed constant after a sufficiently large value of K. The median bias-reducing adjustment $A^\dagger(\theta)$ has tth element

$$A_t^\dagger(\theta) = \frac{1}{2}\mathrm{tr}\big[\{i(\theta)\}^{-1}(P_t(\theta) + Q_t(\theta))\big] - \{i_t(\theta)\}^{\mathrm{T}} K^\dagger(\theta). \tag{4}$$

The quantities $P_t(\theta) = E_\theta[s(\theta)s^{\mathrm{T}}(\theta)s_t(\theta)]$ and $Q_t(\theta) = E_\theta[-j(\theta)s_t(\theta)]$ in equation (4) are those introduced by Kosmidis and Firth[14] for mean bias-reduction, and $K^\dagger(\theta)$ is a $(p+1)$-vector with tth element $K_t^\dagger(\theta) = \{i^t(\theta)\}^{\mathrm{T}} K_t(\theta)$, where $K_t(\theta)$ is another $(p+1)$-vector with uth element

$$K_{tu}(\theta) = \mathrm{tr}\left[\frac{i^t(\theta)\{i^t(\theta)\}^{\mathrm{T}}}{i^{tt}(\theta)}\left(\frac{1}{3}P_u(\theta) + \frac{1}{2}Q_u(\theta)\right)\right]$$

In the context of meta-regression values of t and u in $\{1, \ldots, p\}$ correspond to the elements of parameter β, and $t, u = p+1$ correspond to parameter ψ. Given that $A^\dagger(\theta)$ is of order $O(1)$, $\hat{\theta}^\dagger$ has the same asymptotic distribution as $\hat{\theta}$,[7] i.e. multivariate normal with mean θ and variance-covariance matrix $\{i(\theta)\}^{-1}$, which can be consistently estimated with $\{i(\hat{\theta}^\dagger)\}^{-1}$.

After some algebra (see Appendix 1 for details), the median bias-reducing adjustment for the random-effects meta-analysis and meta-regression models has the form

$$A^\dagger(\theta) = \begin{pmatrix} 0_p \\ \frac{1}{2}\mathrm{tr}[W(\psi)H(\psi)] + \frac{1}{3}\frac{\mathrm{tr}[W(\psi)^3]}{\mathrm{tr}[W(\psi)^2]} \end{pmatrix} \tag{5}$$

where $H(\psi) = X(X^{\mathrm{T}} W(\psi) X)^{-1} X^{\mathrm{T}} W(\psi)$. Substituting equation (5) in the expression for $s^\dagger(\theta)$ gives that the median bias-reducing adjusted score functions for β and ψ are $s_\beta^\dagger(\theta) = s_\beta(\theta)$ and

$$s_\psi^\dagger(\theta) = s_\psi(\theta) + \frac{1}{2}\mathrm{tr}[W(\psi)H(\psi)] + \frac{1}{3}\frac{\mathrm{tr}[W(\psi)^3]}{\mathrm{tr}[W(\psi)^2]}$$

respectively.

4.2 Computation of median bias-reduced estimator

A direct approach for computing the estimator $\hat{\theta}^\dagger = (\hat{\beta}^{\dagger\mathrm{T}}, \hat{\psi}^\dagger)^\mathrm{T}$ is through a modification of the two-step iterative process in Kosmidis et al.[3] At the jth iteration ($j = 1, 2, \ldots$)

(1) Calculate $\beta^{(j)}$ by weighted least squares as $\beta^{(j)} = (X^\mathrm{T} W(\psi^{(j-1)}) X)^{-1} X^\mathrm{T} W(\psi^{(j-1)}) y$
(2) Solve $s_\psi^\dagger(\theta^{(j)}(\psi)) = 0$ with respect to ψ, where $\theta^{(j)}(\psi) = (\beta^{(j)\mathrm{T}}, \psi)^\mathrm{T}$.

In the above steps, $\beta^{(j)}$ is the candidate value for $\hat{\beta}^\dagger$ at the jth iteration and $\psi^{(j-1)}$ is the candidate value for $\hat{\psi}^\dagger$ at the $(j-1)$th iteration. The equation in step 2 is solved numerically, by searching for the root of the function $s_\psi^\dagger(\beta^{(j)}, \psi)$ in a predefined positive interval. For the computations in this manuscript, we use the DerSimonian and Laird[1] estimate of ψ as starting value $\psi^{(0)}$. The iterative process is then repeated until the components of the score function $s^\dagger(\theta)$ are all less than $\epsilon = 1 \times 10^{-6}$ in absolute value at the current estimates.

4.3 Median bias-reducing penalised likelihood

Although it is not generally true that $s^\dagger(\theta)$ is the gradient of a suitable penalised log-likelihood, in this case $s^\dagger(\theta)$ is the gradient of the median BRPL

$$l^\dagger(\theta) = l(\theta) - \frac{1}{2} \log |X^\mathrm{T} W(\psi) X| - \frac{1}{6} \log[\mathrm{tr}(W(\psi)^2)] \tag{6}$$

Hence, $\hat{\theta}^\dagger$ is also the maximum median BRPL estimator. The median BRPL in equation (6) differs from the mean BRPL derived in Kosmidis et al.[3] by the term $-\log[\mathrm{tr}(W(\psi)^2)]/6$.

An advantage of the median BRPL estimators over mean BRPL ones is that the former are equivariant under monotone component-wise parameter transformations.[7] In the context of random-effects meta-analysis and meta-regression, this equivariance implies that not only we get a median bias-reduced estimator of ψ, but we also get median bias-reduced estimates of the standard errors for β by calculating the square roots of the diagonal elements of $\{i(\theta)\}^{-1}$ in equation (3) at $\psi^\dagger$. This is because $i(\theta)$ is a function of ψ only, and moreover the square roots of the diagonal elements of $\{i(\theta)\}^{-1}$ are monotone functions of ψ.

4.4 Penalised likelihood-based inference

For inference about either the components of the fixed-effect parameters β or the between-study heterogeneity ψ, we propose the use of the median BRPL ratio. If $\theta = (\tau^\mathrm{T}, \lambda^\mathrm{T})^\mathrm{T}$ and $\hat{\lambda}_\tau^\dagger$ is the maximiser of $l^\dagger(\theta)$ for fixed τ, then the same arguments as in Kosmidis et al.[3] can be used to show that the logarithm of the median BRPL ratio statistic

$$2\{l^\dagger(\hat{\tau}^\dagger, \hat{\lambda}^\dagger) - l^\dagger(\tau, \hat{\lambda}_\tau^\dagger)\} \tag{7}$$

has a $\chi^2_{\dim(\tau)}$ asymptotic distribution, as K goes to infinity. Specifically, the adjustment to the score function is additive and of order $O(1)$. As a result, the extra terms in the asymptotic expansion of the logarithm of the median BRPL that depend on the penalty and its derivatives disappear as information increases, and the expansion has the same leading term as that of the log-likelihood (see, for example, Pace and Salvan[15] Section 9.4).

5 Cocoa intake and blood pressure reduction data (revisited)

The ML estimate, the maximum mean BRPL estimate and the maximum median BRPL estimate of the heterogeneity parameter in the meta-analysis model in Section 2 are $\hat{\psi} = 4.199$, $\hat{\psi}^* = 5.546$, and $\hat{\psi}^\dagger = 6.897$, respectively. The estimates of the common effect are $\hat{\beta} = -2.799$, $\hat{\beta}^* = -2.811$, and $\hat{\beta}^\dagger = -2.818$, with standard errors 1.000, 1.128, and 1.242, respectively. The bias-reduced estimates of ψ and, as a consequence, the corresponding estimated standard errors for β are larger than their ML counterparts, which is typical in random-effects meta-analysis. The iterative process used for computing the ML, maximum mean BRPL, and maximum median BRPL estimates converged in 4, 5, and 11 iterations, respectively. The computational run-time for the two-step iterative process which computes the ML, maximum mean BRPL, and maximum median BRPL estimates is 1.1×10^{-2}, 1.8×10^{-2}, and 1.1×10^{-2} seconds, respectively.
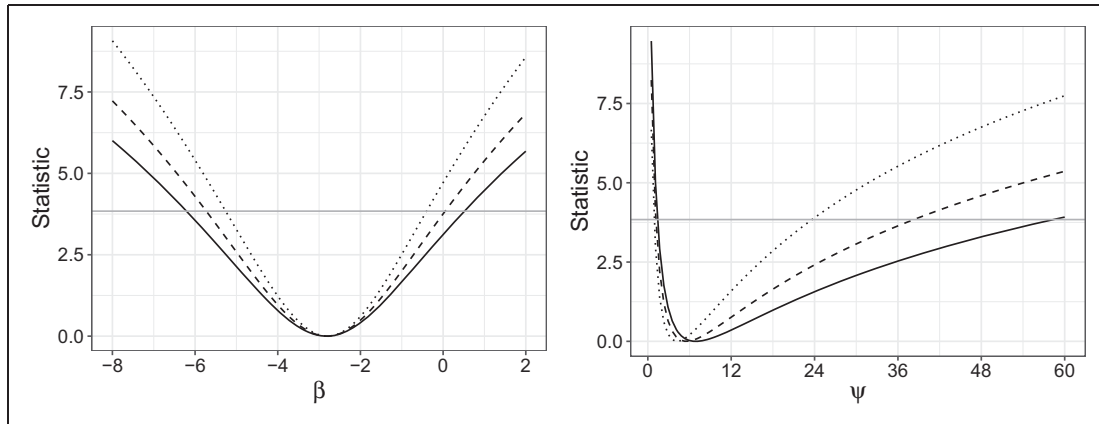
Figure 2. Plot of LR (dotted), mean BRPL (dashed) and median BRPL (solid) ratio statistic in equation (7) when τ is β (left) and ψ (right). The horizontal line is the 95% quantile of the limiting χ_1^2 distribution, and its intersection with the values of the statistics results in the endpoints of the corresponding 95% confidence intervals.
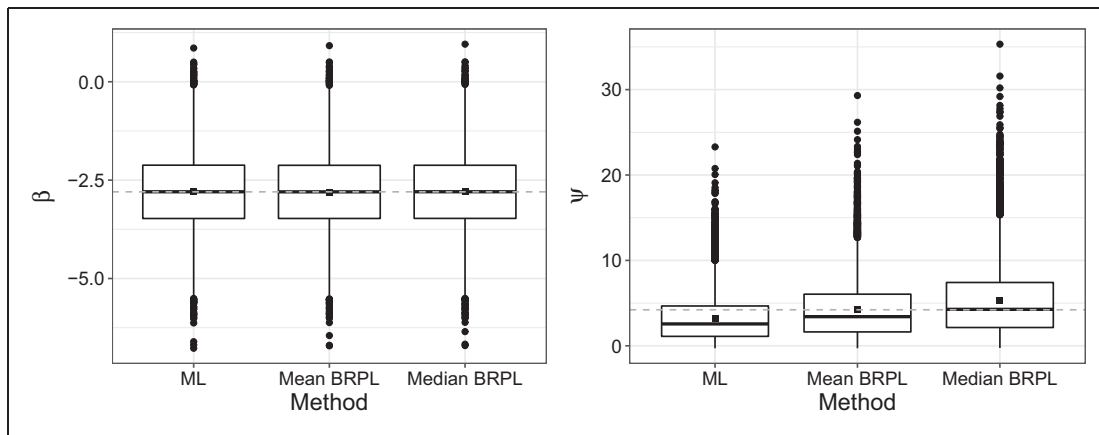
Figure 3. Boxplots for the ML, the maximum mean BRPL, and the maximum median BRPL estimates of β and ψ as calculated from 10,000 simulated samples under the ML fit using the cocoa data.[8,10] The square point is the empirical mean of the estimates. The dashed grey horizontal line is at the parameter value used to generate the data.

Figure 2 shows the value of LR, mean BRPL and median BRPL ratio statistic in equation (7) for a range of values of τ, when τ is either β or ψ. Here and in the following simulation studies, we compare median BRPL ratio statistic with only LR and mean BRPL ratio statistics because the mean BRPL ratio statistic is a strong competitor against other alternatives in terms of inferential performance.[3] The horizontal line in Figure 2 is the 95% quantile of the limiting χ_1^2 distribution, and its intersection with the values of the statistics results in the endpoints of the corresponding 95% confidence intervals. For both β and ψ, the confidence intervals based on the LR statistic are the narrowest and the confidence intervals based on the median BRPL ratio statistic are the widest. Specifically, the 95% confidence intervals for β are $(-6.21, 0.52)$, $(-5.73, 0.05)$, and $(-5.26, -0.40)$ for the median BRPL ratio statistic, mean BRPL ratio statistic, and LR statistic, respectively. The corresponding 95% confidence intervals for ψ are $(1.4, 58.0)$, $(1.0, 38.5)$, and $(1.1, 23.5)$, respectively. Contrary to the LR test, the mean BRPL and median BRPL ratio tests suggest that there is only weak evidence that cocoa consumption affects diastolic blood pressure with p-values of 0.053 and 0.077.

In order to further investigate the performance of the three approaches to estimation and inference, we performed a simulation study where we simulated 10,000 independent samples from the random-effects meta-analysis model with parameter values set to the ML estimates reported earlier, i.e. $\beta_0 = -2.799$ and $\psi_0 = 4.199$. Figure 3 shows boxplots of the estimates of β and ψ calculated from each of the 10,000 simulated samples. The distributions of the three alternative estimators for β are similar. On the other hand, the ML estimator of ψ has a

Table 1. Empirical *p*-value distribution (%) for the tests based on the LR statistic, the mean BRPL ratio statistic, and the median BRPL ratio statistic in the cocoa data[8,10] setting.

$\alpha \times 100$	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
LR	5.7	8.4	11.8	18.2	34.6	57.8	79.2	91.8	96.1	98.0	99.3
Mean BRPL ratio	1.6	3.7	6.7	12.2	28.4	52.8	76.6	90.9	95.5	97.9	99.1
Median BRPL ratio	0.6	1.8	4.1	8.6	23.1	48.5	74.2	90.0	95.0	97.5	99.1

large negative mean bias, maximum median BRPL tends to over-correct for that bias, while maximum mean BRPL almost fully corrects for the bias of ML estimator. The distribution of the median BRPL estimates has the heaviest right tail. The simulation-based estimates of the probabilities of underestimation for ψ, $P_{\psi_0}(\hat{\psi} \leq \psi_0)$, $P_{\psi_0}(\hat{\psi}^* \leq \psi_0)$ and $P_{\psi_0}(\hat{\psi}^\dagger \leq \psi_0)$ are 0.708, 0.591, and 0.493 for the ML, maximum mean BRPL, and maximum median BRPL, respectively, illustrating how effective maximising the median BRPL in equation (6) is in reducing the median bias of the maximum likelihood estimator of ψ.

The simulated samples were also used to calculate the empirical *p*-value distribution for the two-sided tests that each parameter is equal to the true values based on the LR statistic, the mean BRPL ratio statistic, and the median BRPL ratio statistic. Table 1 shows that the empirical *p*-value distribution for the mean and median BRPL ratio statistics are closest to uniformity, with the latter being slightly more conservative than the former. The coverage probability of the 95% confidence intervals of β based on the mean BRPL ratio and the median BRPL ratio are notably closer to the nominal level than those based on the likelihood ratio. Specifically, the coverage probabilities for β are 88%, 93%, and 96% for LR, mean BRPL ratio, and median BRPL ratio, respectively, and the corresponding coverage probabilities for ψ are 88%, 94%, and 96%, respectively.

6 Simulation study

More extensive simulations under the random-effects meta-analysis model (1) are performed here using the design in Brockwell and Gordon.[16] Specifically, the data y_i, $i \in \{1, \ldots, K\}$ are simulated from model (1) with true fixed-effect parameter $\beta = 0.5$. The within-study variances $\hat{\sigma}_i^2$ are independently generated from a χ_1^2 distribution and are multiplied by 0.25 before restricted to the interval $(0.009, 0.6)$. Eleven values of the between-study variance ψ ranging from 0 to 0.1 are chosen, and the number of studies K ranges from 5 to 200. For each combination of ψ and K considered, we simulated 10,000 data sets initialising the random number generator at a common state. The within-study variances were generated only once and kept fixed while generating the samples.

Zeng and Lin[6] compared the performance of their proposed double resampling method with the DerSimonian and Laird[1] method, the profile likelihood method in Hardy and Thompson,[17] and the resampling method in Jackson and Bowden[5] and showed that the double resampling method improves the accuracy of statistical inference. Based on these results, Kosmidis et al.[3] compared the performance of their mean BRPL approach with the double resampling method and illustrated that the former results in confidence intervals with coverage probabilities closer to the nominal level than the alternative methods.

We take advantage of the results reported in Zeng and Lin[6] and Kosmidis et al.[3] and evaluate the performance of estimation and inference based only on the median BRPL with that based on the likelihood and the mean BRPL. The estimators of the fixed and random-effect parameters obtained from the three methods are calculated using variants of the two-step algorithm described in Section 4.2. In the second step of the algorithm, the candidate values for the ML, and maximum mean and median BRPL estimators of the between-study variance ψ are calculated by searching for the root of the partial derivatives of $l(\theta)$, $l^*(\theta)$, and $l^\dagger(\theta)$ with respect to ψ, in the interval (0, 3).

First, we compare the performance of the ML, maximum mean BRPL and maximum median BRPL estimators in terms of percentage of underestimation. Figure 4 shows that the median bias-reducing adjustment is the most effective in reducing median bias even for small values of K. As expected, the ML and maximum mean BRPL estimators also approach the limit of 50% underestimation as K grows, with the latter being closer to 50% than the former. Figure 5 shows that maximum median BRPL is also effective in reducing the mean bias of the ML estimator of ψ but only for moderate to large values of K, while maximum mean BRPL results in estimators with the smallest bias.

Figures 6 and 7 show the estimated coverage probability for the one-sided and two-sided confidence intervals for β based on the LR, mean BRPL ratio and median BRPL ratio statistics at the 95% nominal level. Figure 8
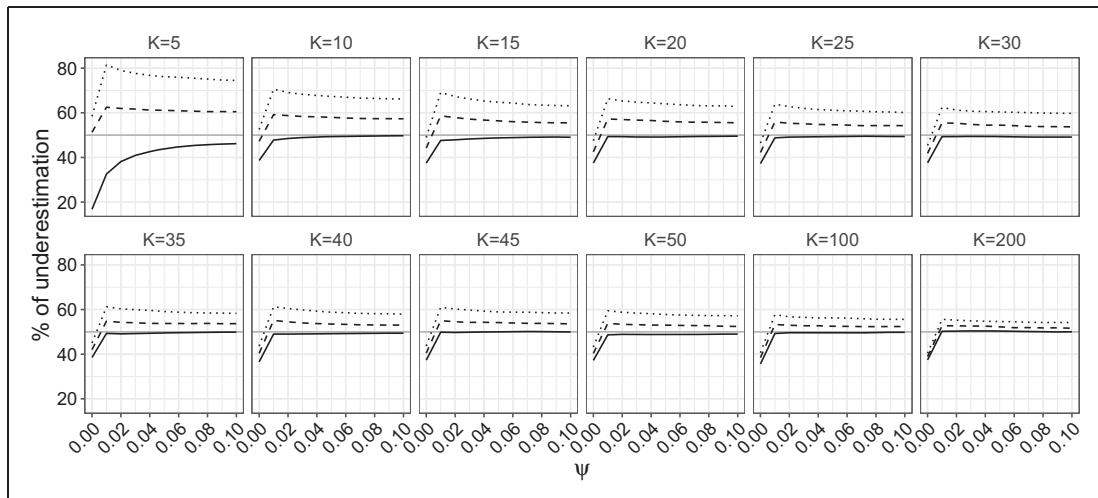
Figure 4. Empirical percentage of underestimation for ψ for random-effects meta-analysis. The percentage of underestimation is calculated for $K \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200\}$ and for increasing values of ψ in the interval $[0, 0.1]$. The curves correspond to the maximum median BRPL (solid), maximum mean BRPL (dashed), and ML (dotted) estimators. The grey horizontal line is at the target of 50% underestimation.
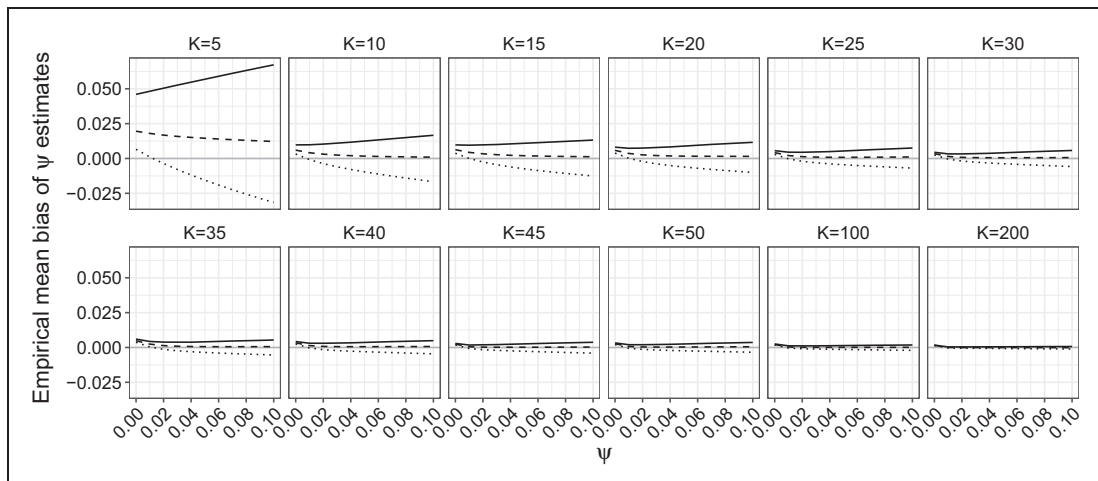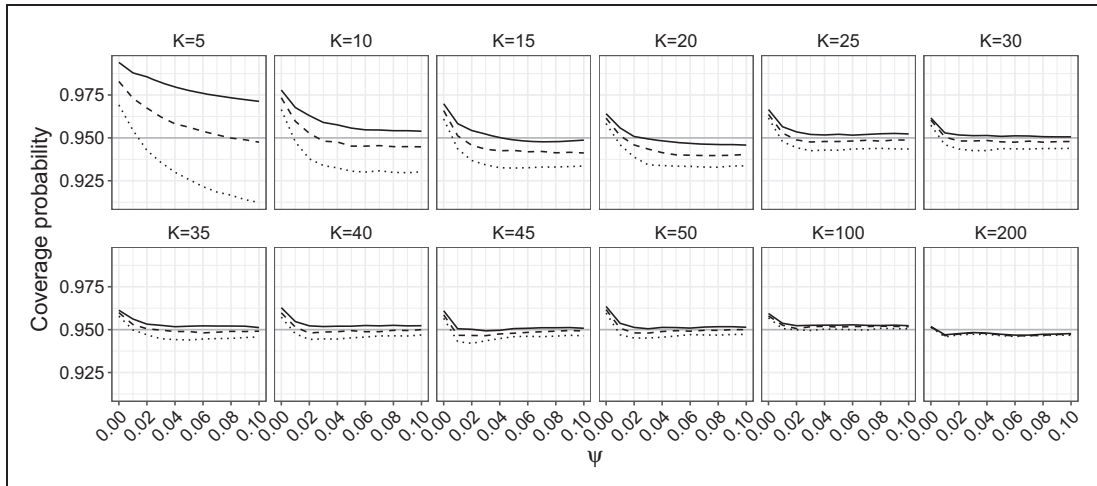

Figure 5. Empirical mean bias of ψ estimates for random-effects meta-analysis. The mean bias is calculated for $K \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200\}$ and for increasing values of ψ in the interval $[0, 0.1]$. The curves correspond to the maximum median BRPL (solid), maximum mean BRPL (dashed), and ML (dotted) estimators. The grey horizontal line is at zero.

shows the estimated coverage probability for the two-sided confidence intervals for ψ based on the LR, mean BRPL ratio and median BRPL ratio statistics at the 95% nominal level. For small values of ψ or small and moderate number of studies K, the empirical coverage of the intervals is larger than the nominal 95% level. In general, the confidence intervals based on mean and median BRPL ratio have empirical coverage that is closer to the nominal level with the latter having generally better coverage. The differences between the three methods diminish as the number of studies K increases.

Figures 9 and 10 give the power of the LR, the mean BRPL ratio, and the median BRPL ratio tests for testing the null hypothesis $\beta = 0.5$ against various alternatives. Specifically, we simulated 10,000 data sets under the alternative hypothesis that parameter β is equal to $b = 0.5 + \delta K^{-1/2}$, where δ ranges from 0 to 2.25. In Figure 9 the power is calculated using critical values of the asymptotic null χ_1^2 distribution of the statistics. In Figure 10, the power is calculated using critical values based on the exact null distribution of each statistic, obtained by simulation under the null hypothesis. In this way, the three tests are calibrated to have size 5%.
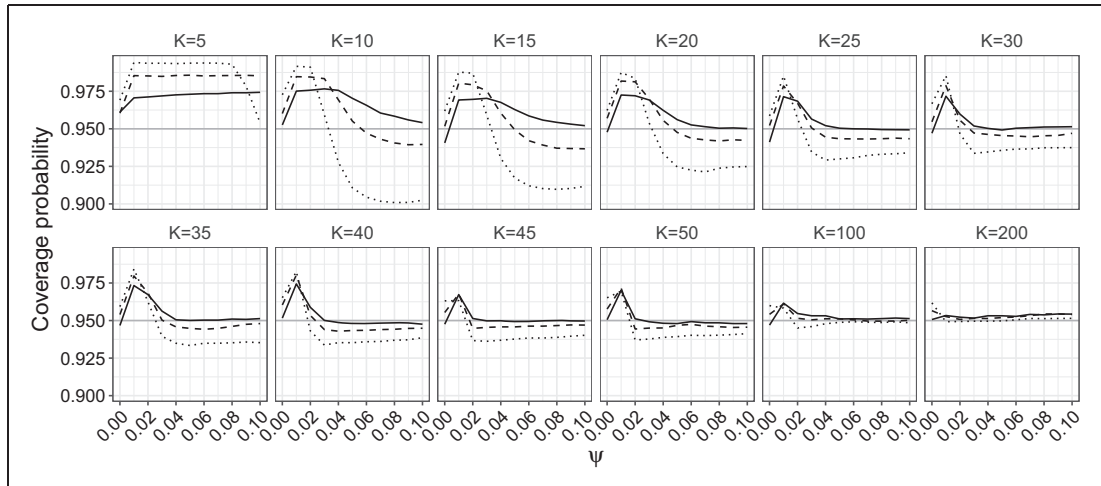
Figure 6. Empirical coverage probabilities of one-sided (right) confidence intervals for β for random-effects meta-analysis. The empirical coverage is calculated for increasing values of ψ in the interval $[0, 0.1]$ and for $K \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200\}$. The curves correspond to nominally 95% internals based on the median BRPL ratio (solid), the mean BRPL ratio (dashed), and the LR (dotted). The grey horizontal line is at the 95% nominal level.
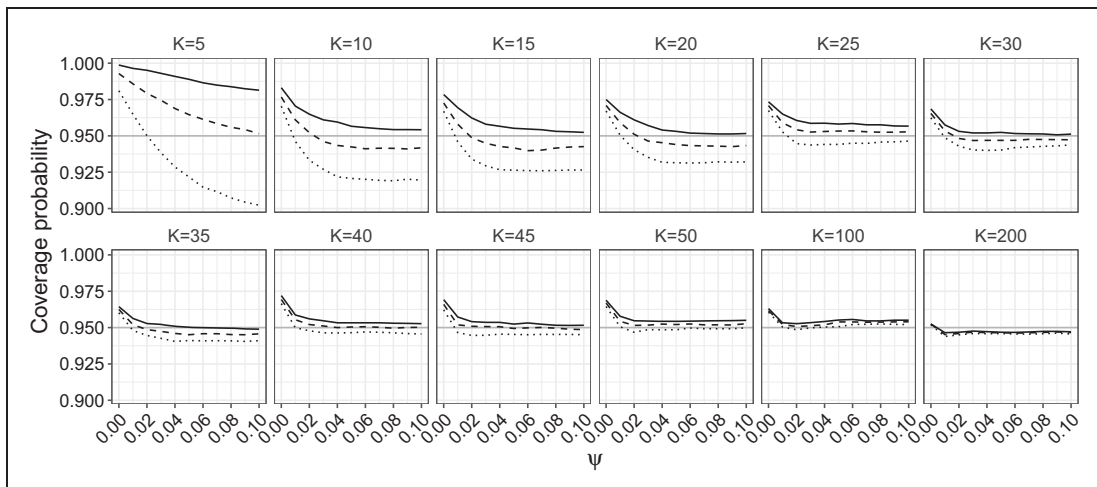

Figure 7. Empirical coverage probabilities of two-sided confidence intervals for β for random-effects meta-analysis. The empirical coverage is calculated for increasing values of ψ in the interval $[0, 0.1]$ and for $K \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200\}$. The curves correspond to nominally 95% confidence intervals based on the median BRPL ratio (solid), the mean BRPL ratio (dashed), and the LR (dotted). The grey horizontal line is at the 95% nominal level.

Figure 9 shows that the three tests have monotone power and for small values of K the LR test yields the largest power. This is because the LR test is oversized, while the mean and median BRPL ratio tests are slightly more conservative and this conservativeness comes at the cost of lower power. As the number of studies K increases, the three tests approach the nominal size and provide similar power. The use of the exact critical values in Figure 10 allows us to compare the performance of the tests without letting the oversizing or the conservativeness of a test skew the power results. Figure 10 shows that the power of the median BRPL ratio test is almost identical to that of the mean BRPL ratio test, and both tests have larger power than the LR test. Again, inference based on either of the two penalised likelihoods becomes indistinguishable from classical likelihood inference as the number of studies increases.

Across all ψ and K values considered, the average number of iterations taken per fit for the two-step iterative process to converge is 6.20, 5.75, and 5.86 iterations for ML, maximum mean BRPL, and maximum median

Figure 8. Empirical coverage probabilities of two-sided confidence intervals for ψ for random-effects meta-analysis. The empirical coverage is calculated with $\beta = 0.5$ and for increasing values of ψ in the interval $[0, 0.1]$ and for $K \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200\}$. The curves correspond to nominally 95% confidence intervals based on the median BRPL ratio (solid), the mean BRPL ratio (dashed), and the LR (dotted). The grey horizontal line is at the 95% nominal level.
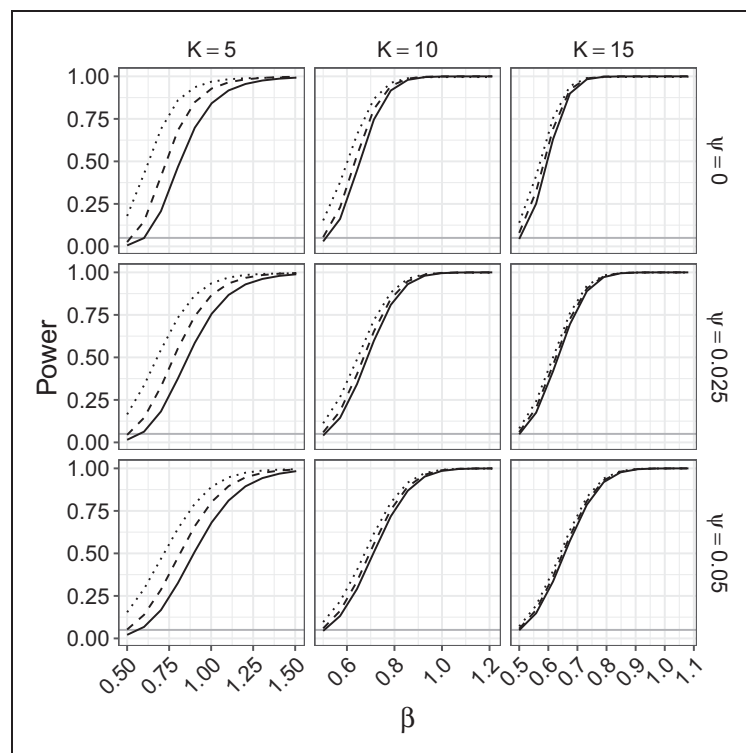
Figure 9. Empirical power of the likelihood-based tests of asymptotic level 0.05 for random-effects meta-analysis for testing $\beta = 0.5$. The empirical power is calculated for increasing values of β, for $K \in \{5, 10, 15\}$ and $\psi \in \{0, 0.025, 0.05\}$. The curves correspond to median BRPL ratio (solid), mean BRPL ratio (dashed), and LR (dotted) tests. The grey horizontal line is at the 5% nominal size.

BRPL, respectively. The average computational run-times for ML, maximum mean BRPL, and maximum median BRPL are 0.005 s, 0.021 s, and 0.017 s, respectively. Figures 1 and 2 in the Supplementary material show the average number of iterations and the average computational run-time taken per fit for the two-step iterative process to converge for each value of K and ψ used in the simulation study. The results show that in all cases
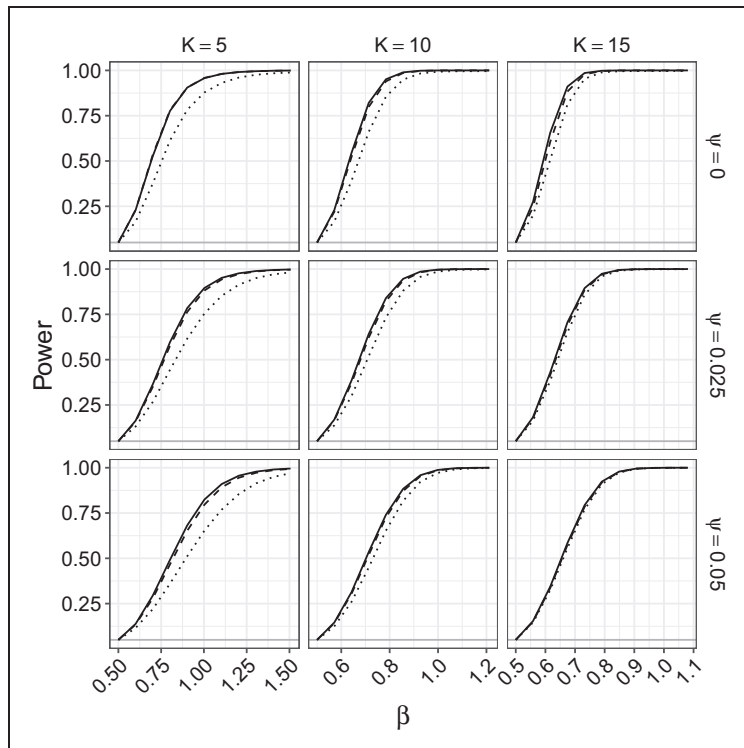
Figure 10. Empirical power of the likelihood-based tests of exact level 0.05 for random-effects meta-analysis for testing $\beta = 0.5$. The empirical power is calculated for increasing values of β, for $K \in \{5, 10, 15\}$ and $\psi \in \{0, 0.025, 0.05\}$. The curves correspond to median BRPL ratio (solid), mean BRPL ratio (dashed), and LR (dotted) tests. The grey horizontal line is at the 5% nominal size.

estimation is achieved rapidly and after a small number of iterations for all three methods, with only negligible overhead with the two bias reducing methods.
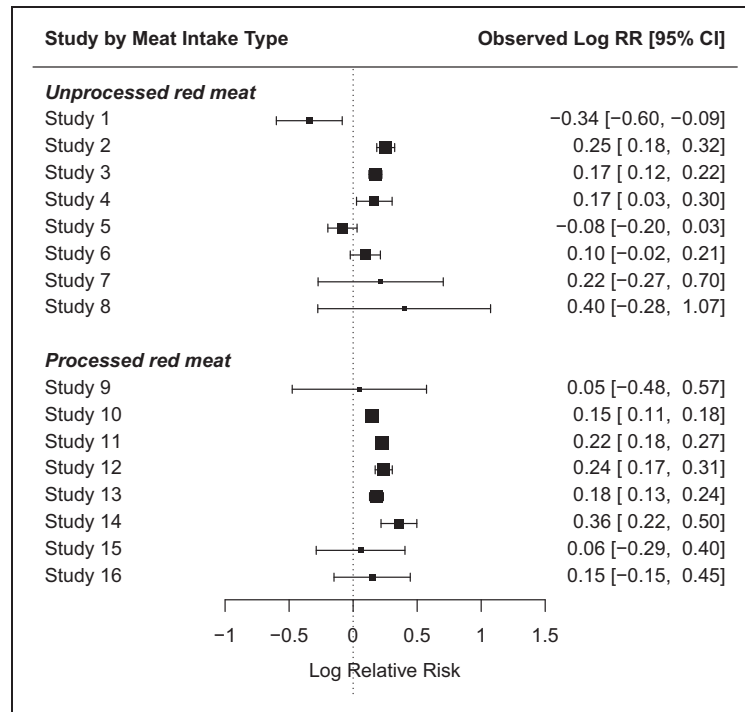
7 Meat consumption data

Larsson and Orsini[9] investigate the association between meat consumption and relative risk of all-cause mortality. The data consists of 16 prospective studies, eight of which are about unprocessed red meat consumption and eight about processed meat consumption. Figure 11 displays the information provided by each study in the meta-analysis. The results from the studies point towards the conclusion that high consumption of red meat, in particular processed red meat, is associated with higher all-cause mortality.

We consider the random-effects meta-regression model, assuming that Y_i has a $N(\beta_0 + \beta_1 x_i, \hat{\sigma}_i^2 + \psi)$, where Y_i is the random variable representing the logarithm of the relative risk reported in the ith study, and x_i takes value 1 if the consumption in the ith study is about processed red meat and 0 if it is about unprocessed meat ($i = 1, \ldots, 16$). Table 2 gives the ML estimates, the mean BRPL estimates, and the median BRPL estimates of β_0, β_1 and ψ, along with the corresponding standard errors for β_0 and β_1. The median BRPL estimate of ψ and the standard errors of the fixed-effect parameters are the largest. The iterative process for computing the ML, maximum mean BRPL, and maximum median BRPL estimates converged in eight, nine, and twelve iterations, in 1.2×10^{-2}, 2.4×10^{-2}, and 1.5×10^{-2} seconds, respectively.

The LR test indicates some evidence for a higher risk associated to the consumption of red processed meat with a p-value of 0.047. On the other hand, the mean and median BRPL ratio tests suggest that there is weaker evidence for higher risk, with p-values of 0.066 and 0.074, respectively. Skovgaard's test also gives weak evidence for higher risk with p-value 0.073.

Similar to Section 5, we performed a simulation study in order to further investigate the performance of the three methods in a meta-regression context. We simulated 10,000 independent samples from the meta-regression model at the ML estimates reported in Table 2. Figure 12 shows boxplots of the estimates of β_0, β_1, and ψ. Maximum likelihood underestimates the parameter ψ, while mean BRPL and median BRPL almost fully

Figure 11. The meat consumption data.[9] Outcomes from 16 studies are reported in terms of the logarithm of the relative risk (Log RR) of all-cause mortality for the highest versus lowest category of unprocessed red meat, and processed meat consumption. Squares represent the mean effect estimate for each study; the size of the square reflects the weight that the corresponding study exerts in the meta-analysis. Horizontal lines represent 95% Wald-type confidence intervals (CI) of the effect estimate of individual studies.

Table 2. ML, maximum mean BRPL, and maximum median BRPL estimates of the model parameters for the meat consumption data.[9]

| Method	β_0	β_1	ψ
ML	0.099 (0.044)	0.106 (0.061)	0.009
	[−0.004,0.189]	[−0.022,0.244]	[0.003,0.030]
Maximum mean BRPL	0.095 (0.050)	0.110 (0.069)	0.012
	[−0.020,0.199]	[−0.040,0.264]	[0.003,0.042]
Maximum median BRPL	0.093 (0.052)	0.111 (0.072)	0.013
	[−0.027,0.203]	[−0.048,0.271]	[0.004,0.048]

Note: Standard errors are reported in parentheses. The 95% confidence intervals based on the LR, mean BRPL ratio and median BRPL ratio are reported in squared brackets.

compensate for the negative bias of ML estimates, with the latter having a slightly heavier right tail. The percentages of underestimation are 72.6%, 56.6%, and 49.9% for the ML, maximum mean BRPL, and maximum median BRPL estimators, respectively.

The simulated samples were also used to calculate the empirical *p*-value distribution for the tests based on the likelihood, mean BRPL and median BRPL ratio statistics. Table 3 shows that the empirical *p*-value distribution for the median BRPL ratio statistic is the one closest to uniformity.

8 Concluding remarks

In this paper we derive the adjusted score equations for the median bias reduction of the ML estimator for random-effects meta-analysis and meta-regression models and describe the associated inferential procedures.
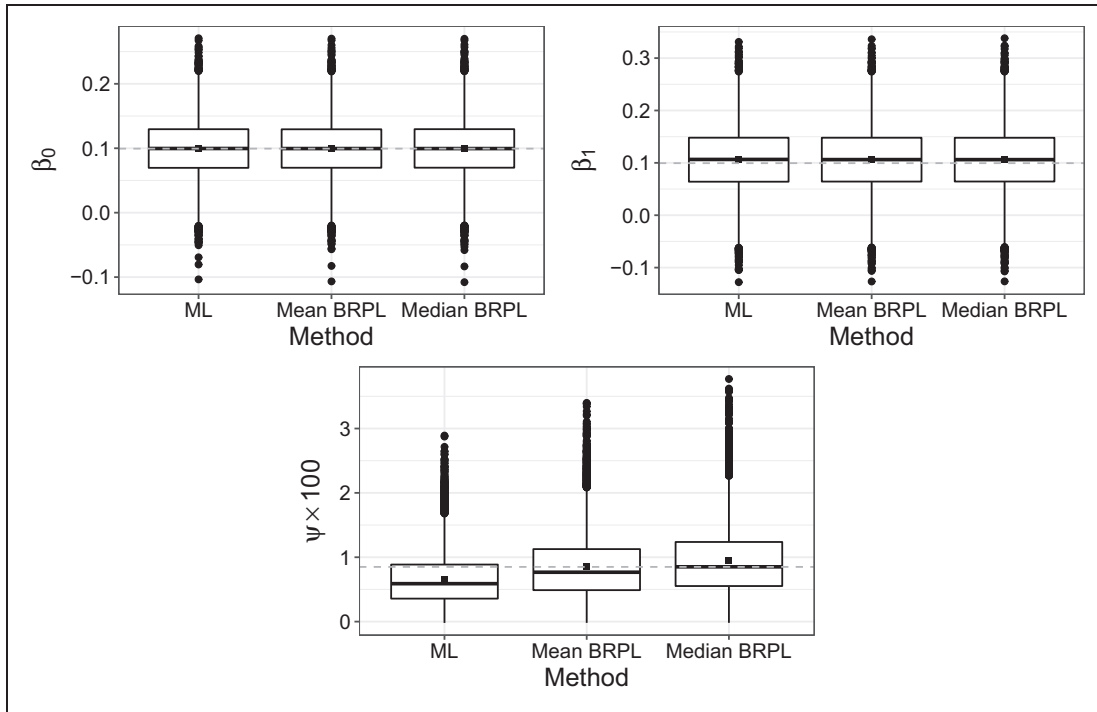
Figure 12. Boxplots for the ML, maximum mean BRPL, and maximum median BRPL estimates of β_0, β_1, and ψ as calculated from 10,000 simulated samples under the ML fit using the meat consumption data.[9] The square point is the mean of the estimates obtained from each method. The dashed grey horizontal line is at the parameter value used to generate the data.

Table 3. Empirical p-value distribution (%) for the tests based on the LR statistic, the mean BRPL ratio statistic, and the median BRPL ratio statistic using the meat consumption data.[9]

$\alpha \times 100$	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
LR	2.2	4.5	7.7	13.1	28.0	50.0	71.7	86.6	92.1	95.3	97.7
Mean BRPL ratio	1.3	3.0	5.6	11.1	25.9	49.8	73.8	89.0	94.2	96.9	98.6
Median BRPL ratio	1.0	2.5	4.9	9.9	25.1	49.7	74.7	89.8	94.8	97.5	98.9

We show that the solution of the median bias-reducing adjusted score equations is equivalent to maximising a penalised log-likelihood. The logarithm of that penalised likelihood differs from the logarithm of the mean BRPL in Kosmidis et al.[3] by a simple additive term. The computation of the maximum median BRPL estimators can be performed through a two-step iteration that involves a weighted least squares update and the solution of a nonlinear equation with respect to a scalar parameter, and which converges rapidly, as illustrated by the computational times and number of iterations reported in the paper. The reported times and number of iterations were computed using a workstation with 24 cores at 2.90 GHz and 80 GB memory running under the CentOS 7 operating system, using one core per data set.

Using various settings we were able to retrieve enough information on the performance of the maximum median BRPL estimators. All our simulation studies illustrate that use of the median BRPL succeeds in achieving median centering in estimation, and results in confidence intervals with good coverage properties. Furthermore, while tests based on the LR suffer from size distortions, the median BRPL ratio statistic results in tests with size and power properties, sometimes better to those of the mean BRPL ratio statistic in Kosmidis et al.[3]

The main advantage of the maximum median BRPL estimators from the maximum mean BRPL ones is their equivariance under monotone component-wise parameter transformations, which, in the case of random-effects meta-regression, leads to median bias-reduced standard errors.

As random-effect models are widely used in practice, the median BRPL method is likely to be useful in models with more complex random-effect structures, such as linear mixed models.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

ORCID iD

Sophia Kyriakou http://orcid.org/0000-0001-8077-6439

References

1. DerSimonian R and Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.
2. Guolo A and Varin C. Random-effects meta-analysis: the number of studies matters. *Stat Meth Med Res* 2017; **26**: 1500–1518.
3. Kosmidis I, Guolo A and Varin C. Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression. *Biometrika* 2017; **104**: 489–496.
4. Van Houwelingen HC, Arends LR and Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; **21**: 589–624.
5. Jackson D and Bowden J. A re-evaluation of the "quantile approximation method" for random effects meta-analysis. *Stat Med* 2009; **28**: 338–348.
6. Zeng D and Lin DY. On random-effects meta-analysis. *Biometrika* 2015; **102**: 281–294.
7. Kenne Pagui EC, Salvan A and Sartori N. Median bias reduction of maximum likelihood estimates. *Biometrika* 2017; **104**: 923–938.
8. Taubert D, Roesen R and Schömig E. Effect of cocoa and tea intake on blood pressure: a meta-analysis. *Arch Intern Med* 2007; **167**: 626–634.
9. Larsson SC and Orsini N. Red meat and processed meat consumption and all-cause mortality: a meta-analysis. *Am J Epidemiol* 2014; **179**: 282–289.
10. Bellio R and Guolo A. Integrated likelihood inference in small sample meta-analysis for continuous outcomes. *Scand J Stat* 2016; **43**: 191–201.
11. Knapp G and Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; **22**: 2693–2710.
12. Huizenga HM, Visser i and Dolan CV. Testing overall and moderator effects in random effects meta-regression. *Br J Math Stat Psychol* 2011; **64**: 1–19.
13. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; **80**: 27–38.
14. Kosmidis I and Firth D. Bias reduction in exponential family nonlinear models. *Biometrika* 2009; **96**: 793–804.
15. Pace L and Salvan A. *Principles of statistical inference: from a neo-Fisherian perspective*. London: World Scientific, 1997.

16. Brockwell SE and Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20**: 825–840.
17. Hardy RJ and Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med* 1996; **15**: 619–629.

Appendix 1

The observed information matrix for the random-effects meta-regression model (1) is

$$
j(\theta) = \begin{pmatrix} X^{\mathrm{T}} W(\psi) X & X^{\mathrm{T}} W(\psi)^2 R(\beta) \\ X^{\mathrm{T}} W(\psi)^2 R(\beta) & R(\beta)^{\mathrm{T}} W(\psi)^3 R(\beta) - \frac{1}{2}\mathrm{tr}[W(\psi)^2] \end{pmatrix}.
$$

For this model

$$
P_t(\theta) = -Q_t(\theta) = \begin{pmatrix} 0_{p\times p} & X^{\mathrm{T}} W(\psi)^2 X_t \\ X^{\mathrm{T}} W(\psi)^2 X_t & 0 \end{pmatrix} \quad (t = 1, \ldots, p),
$$

and

$$
P_{p+1}(\theta) = \begin{pmatrix} X^{\mathrm{T}} W(\psi)^2 X & 0_p \\ 0_p^{\mathrm{T}} & \mathrm{tr}(W(\psi)^3) \end{pmatrix} \quad \text{and} \quad Q_{p+1}(\theta) = \begin{pmatrix} 0_{p\times p} & 0_p \\ 0_p^{\mathrm{T}} & -\mathrm{tr}(W(\psi)^3) \end{pmatrix},
$$

where $0_{p\times p}$ is the $p \times p$ zero matrix and X_t is the tth column of X. The median bias-reducing adjustment for θ is obtained by plugging the above expressions into equation (4). The sum $P_t(\theta) + Q_t(\theta)$ $(t = 1, \ldots, p+1)$ is also given in the Appendix of Kosmidis et al.[3]