Random-effects meta-analysis: The number of studies matters

Annamaria Guolo[1] **and Cristiano Varin**[2]

Abstract

This paper investigates the impact of the number of studies on meta-analysis and meta-regression within the random-effects model framework. It is frequently neglected that inference in random-effects models requires a substantial number of studies included in meta-analysis to guarantee reliable conclusions. Several authors warn about the risk of inaccurate results of the traditional DerSimonian and Laird approach especially in the common case of meta-analysis involving a limited number of studies. This paper presents a selection of likelihood and non-likelihood methods for inference in meta-analysis proposed to overcome the limitations of the DerSimonian and Laird procedure, with focus on the effect of the number of studies. The applicability and the performance of the methods are investigated in terms of Type I error rates and empirical power to detect effects, according to scenarios of practical interest. Simulation studies and applications to real meta-analyses highlight that it is not possible to identify an approach uniformly superior to alternatives. The overall recommendation is to avoid the DerSimonian and Laird method when the number of meta-analysis studies is modest and prefer a more comprehensive procedure that compares alternative inferential approaches. `R` code for meta-analysis according to all the inferential methods examined in the paper is provided.

Keywords: likelihood, meta-analysis, random-effects model, small number of studies, Type I error

[1]University of Padova, Padova, Italy
[2]Ca' Foscari University, Venice, Italy

Corresponding author:
Annamaria Guolo, University of Padova, Via Cesare Battisti 241, I-37121, Padova, Italy. Phone: +39 049 827 4150. E-mail: annamaria.guolo@unipd.it

1 Introduction

Meta-analysis of aggregate data is the most common approach to synthesize information from independent studies about the same issue of interest. Meta-analysis has a long tradition in the medical literature, although its usage extends beyond to reach almost any area of research[1]. In this paper, the attention will be focused on the random-effects meta-analysis formulation, initially proposed by DerSimonian and Laird[2], as a tool to account for between-study variability. Criticisms towards random-effects modelling include the use of study weights that are not inversely proportional to the study sample sizes, difficulty to assess the distributional assumption of the unobserved random-effects and the fact that the model does not provide any explanation of the source of between-study variability. Despite the criticism, the random-effects approach is popular as an effective way to combine studies affected by heterogeneity.

The DerSimonian and Laird[2] procedure provides simple expressions for inference on parameters of the random-effects model. The straightforward application of the procedure and the possibility of easy implementation with any standard software are at the basis of its success. Nevertheless, drawbacks of the procedure have been widely pointed out, including the inaccurate coverage of confidence intervals, smaller than it should be, and the flawed P-value of hypothesis testing, lower than expected[3,4]. Several authors highlight the need of a substantial number of studies included in the meta-analysis in order to guarantee reliable inferential results[5,6]. Unfortunately, such a requirement is typically not satisfied in applications since the number of studies involved in the meta-analysis is often considerably small. Modifications of standard inferential procedures are needed.

This paper presents a selection of methods developed to provide more reliable instruments for meta-analysis with a small number of studies. The described methods are freely available to practitioners through the `R` programming language[7]. All the competing methods are discussed by referring to real meta-analyses from the medical literature. A simulation study is performed to compare the different techniques in terms of Type I error rates and empirical power to detect effects, according to scenarios of practical

interest. In order to facilitate meta-analysis according to different approaches, `R` code is made available as supplementary material and illustrated in the Appendix.

2 The traditional approach to meta-analysis

2.1 Fixed-effects and random-effects models

The main goal of meta-analysis is inference on a true effect β, starting from the information provided by n separate comparable studies on the same issue of interest. Let Y_i denote the measure of β provided by study $i = 1, \ldots, n$, such as, for example, the log-odds ratio or the standardized mean difference and let σ_i^2 denote the within-study variance as a measure of the uncertainty of each study in providing the measure Y_i. A distinction is made between fixed-effects and random-effects modeling in meta-analysis, see Schmidt et al. (2009)[8] and Borenstein et al. (2010)[9] for an extensive discussion. The fixed-effects linear model is specified as

$$Y_i = \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2), \tag{1}$$

where errors ε_i are assumed to be independent Normal variables with zero mean and within-study variance σ_i^2. Model (1) takes account of the sampling error related to the sampling procedure within each study inserted in the meta-analysis and measured by the within-study variance σ_i^2. The random-effects model includes a second source of variability, related to the sampling error caused by variation among the studies. Following the specification by DerSimonian and Laird[2], the linear random-effects model is

$$Y_i = \beta_i + \varepsilon_i, \quad \beta_i = \beta + \eta_i, \tag{2}$$

where β_i and η_i indicate the random-effect and the error accounting for between-study variability, respectively. Commonly, errors η_i are assumed to be independent Normal variables with zero mean and between-study variance τ^2, independently of ε_i. Marginally, Y_i follows the Normal distribution $Y_i \sim \text{Normal}(\beta, \tau^2 + \sigma_i^2)$. A common assumption is

that each study is based on a sample size large enough to consider the within-study variance σ_i^2 as known and equal to the estimate provided by each study. The assumption will be adopted hereafter. Accordingly, the unknown parameter vector is $\boldsymbol{\theta} = (\beta, \tau^2)^\top$.

The selection of a fixed-effects or a random-effects model has been largely discussed in the literature[8–11] and summarized in the Cochrane guidelines[12], given the implications of the choice on the analysis and on the interpretation of the results. The fixed-effects model assumes that there is only one source of variation in the studies and that differences in the observed effects are due to the internal sampling procedure only. No other difference among the studies inserted in the meta-analysis is allowed. The object of inference β is thus the common effect size shared by the studies[9] or, equivalently, the mean effect in the population. The random-effects model includes an additional source of sampling variability due to variation across studies. The between-study variability reflects the effect size differences from study to study occurring as a consequence of patients characteristics and implementation of interventions, among others. Hedges and Vevea[13] highlight that the choice of a model formulation is related to the inferential goal, noting that conclusions from a fixed-effects model refer to the available studies at hand and do not generalize beyond, a result the authors call *conditional inference*. Conversely, the random-effects model produces an *unconditional inference*[13], by allowing to generalize the conclusions beyond the observed studies to other studies with similar, although variable, characteristics. See also Section 9.5.4 *Incorporating heterogeneity into random-effects models* of the Cochrane Handbook for Systematic Reviews of Interventions[12].

2.2 Detecting heterogeneity

The evaluation of the heterogeneity among the studies inserted in a meta-analysis has received a lot of attention in literature, see Hardy and Thompson[14], Whitehead (Chapter 6)[15] and Viechtbauer[16]. Starting from the random-effects model (2), the detection of heterogeneity entails testing the null homogeneity hypothesis $\tau^2 = 0$, which corresponds to the fixed-effects model (1). The estimator of β under the homogeneity

hypothesis is a weighted average of Y_i, with weights given by the inverse of the within-study variances,

$$\hat{\beta}_{\text{FE}} = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}, \quad w_i = \frac{1}{\sigma_i^2}. \tag{3}$$

Commonly, the homogeneity hypothesis is tested via the so-called Q test based on Cochran[17] statistic[18],

$$Q = \sum_{i=1}^{n} (Y_i - \hat{\beta}_{\text{FE}})^2 w_i,$$

which has an asymptotic chi-squared distribution χ^2_{n-1} with $n-1$ degrees of freedom under the null hypothesis. Large values of Q provide evidence against the homogeneity hypothesis. Many authors investigated the performance of the Q test to detect between-study heterogeneity[14,16], showing that the test suffers for very low power when studies have small sample size or the number of studies n is small. In these circumstances, a non-significant Q does not reliably identify lack of heterogeneity. Therefore, many authors recommend the use of the random-effects model despite the result of the Q test.

Alternative procedures assess heterogeneity through diagnostic plots, see Galbraith[19], Hardy and Thompson[14] and Julious and Whitehead[20], or propose to measure the impact of heterogeneity through statistics independent of the number of studies and of the effect measure, see Higgins and Thompson[21] and Higgins et al.[22].

A caveat related to the evaluation of the heterogeneity is that if the number n of studies is small, then the estimation of the between-study variance can be inaccurate, independently of the sample size of each study. Inferential conclusions on the effect β are affected as well in terms of precision, as it will be illustrated in the following section. Recommendations about how to design additional studies to be included in a meta-analysis are discussed in Roloff et al.[23].

2.3 The DerSimonian and Laird approach

The traditional approach to random-effects meta-analysis proposed by DerSimonian and Laird[2] is based on the method of moments. The estimator of β is the weighted average

$$\hat{\beta}_{\mathrm{DL}} = \frac{\sum_{i=1}^{n} \hat{w}_i Y_i}{\sum_{i=1}^{n} \hat{w}_i}, \quad \hat{w}_i = \frac{1}{\hat{\tau}^2_{\mathrm{DL}} + \sigma_i^2}, \tag{4}$$

where $\hat{\tau}^2_{\mathrm{DL}}$ is the estimator of τ^2 defined as

$$\hat{\tau}^2_{\mathrm{DL}} = \max\left\{0, \frac{q-(n-1)}{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / \sum_{i=1}^{n} w_i}\right\}$$

and q is the observed value of the Cochran's Q statistic. The left-censored form of $\hat{\tau}^2_{\mathrm{DL}}$ guarantees the non-negativity of the estimate of τ^2. The estimator $\hat{\beta}_{\mathrm{DL}}$ has variance $\mathrm{var}(\hat{\beta}_{\mathrm{DL}}) = (\sum_{i=1}^{n} w_i)^{-1}$. Inference relies on the central limit theorem which assures that the standardized $\hat{\beta}_{\mathrm{DL}}$ has a standard Normal distribution asymptotically on n.

The estimator $\hat{\beta}_{\mathrm{DL}}$ has the substantial advantage of a simple computation, a feature that, at the time of the proposal in the mid-1980s, made the procedure very attractive. Drawbacks have been pointed out in successive years. The plug-in procedure which substitutes τ^2 with $\hat{\tau}^2_{\mathrm{DL}}$ into the expressions of $\hat{\beta}_{\mathrm{DL}}$ and $\mathrm{var}(\hat{\beta}_{\mathrm{DL}})$ may give rise to unreliable inferential conclusions, because the variability associated to the estimation of τ^2 is not accounted for[3,4]. As a consequence, confidence intervals for β are narrower on average than they should be and P-values associated to hypothesis testing are smaller than expected, especially when n is modest.

Example 1: Local anaesthesia data. Intrauterine pathologies can be identified by ambulatory hysteroscopy, a diagnostic instrument which is becoming prominent as a standard of care. Local anaesthesia can be a useful instrument to control pain associated to the examination, although the efficacy of anaesthesia is controversial. Cooper, Khan and Clark[24] perform a meta-analysis about the efficacy of different anaesthetics for pain control during hysteroscopy. We consider the portion of the data referring to the use of paracervical anaesthesia, consisting of information from five randomized controlled

studies[25–29], two of them based on a double blinded design[26,27]. As reported by Cooper, Khan and Clark[24], the study design was "restricted [...] to randomized controlled studies to minimize selection bias". The outcome is the standardized mean difference of pain scores, measured at the time of procedure and as vasovagal episodes. The forest plot of the data is reported in Figure 1, while details about the studies inserted in the meta-analysis are summarized in Table 4. The inferential interest lies in the significance of the parameter β associated to pain reduction. The estimate $\hat{\beta}_{\mathrm{DL}}$ is -1.28, with standard error $\mathrm{se}(\hat{\beta}_{\mathrm{DL}}) = 0.48$. Standard meta-analysis based on DerSimonian and Laird concludes for a highly evidence of pain reduction, with an associated P-value equal to 0.007 and 95% confidence interval (-2.22, -0.35), see Figure 1.

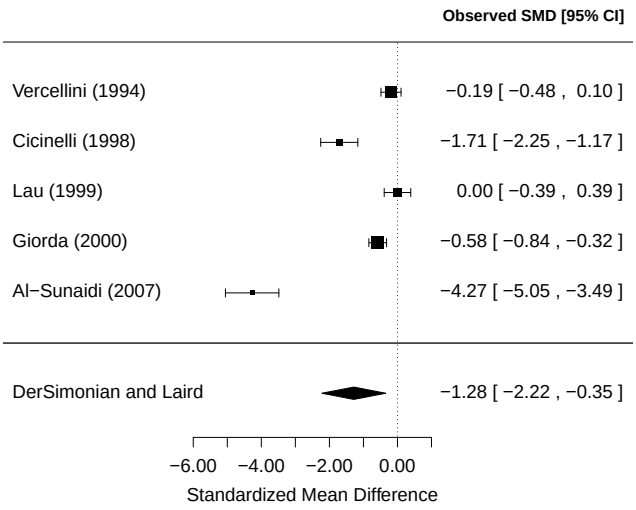

Figure 1: Forest plot of local anaesthesia data[24]. Outcomes from the meta-analysis studies are reported in terms of standardized mean difference (SMD), together with the associated 95% confidence interval in square brackets (95% CI). The DerSimonian and Laird estimate of β is reported together with the 95% confidence interval in square brackets.

In order to evaluate the reliability of the DerSimonian and Laird procedure, we perform a simulation exercise. We generate 1 000 replicates of the local anaesthesia data from model (2) under the null hypothesis $\beta = 0$, with τ^2 equal to $\hat{\tau}^2_{\mathrm{DL}}$ and values of σ_i^2

Study	Treated	Controls
Vercellini (1994)[25]	87	90
Cicinelli (1998)[26]	36	36
Lau (1999)[27]	49	50
Giorda (2000)[28]	121	119
Al-Sunaidi (2007)[29]	42	42

Table 1: Characteristics of the studies included in the local anaesthesia data[24].

and n equal to those of the local anaesthesia data. For each simulated dataset, we compute the standardized estimate $\hat{\beta}_{\text{DL}}/\text{se}(\hat{\beta}_{\text{DL}})$. Figure 2 reports the histogram of the distribution of the simulated standardized estimate, that is expected to follow the standard Normal distribution. The comparison between the sample quantiles of the simulated standardized estimates and the theoretical quantiles from the standard Normal distribution shows that the inaccuracy of the DerSimonian and Laird procedure that does not guarantee the nominal probability of Type I error. Indeed, only 86.8% of the simulated standardized estimates fall inside the 95% theoretical interval (-1.96, 1.96), see the dark grey area of the histogram. In other terms, the simulation indicates that the probability of Type I error is 13.2% instead of the nominal 5%. □

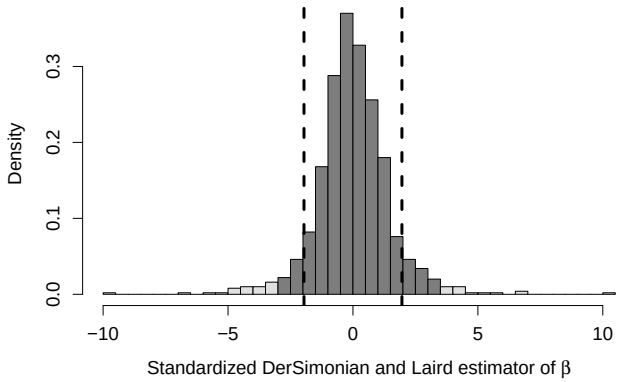

Figure 2: Simulated distribution of the standardized estimator $\hat{\beta}_{\text{DL}}/\text{se}(\hat{\beta}_{\text{DL}})$, based on 1 000 replicates of local anaesthesia data[24] under the null hypothesis $\beta = 0$. Vertical dashed lines identify the theoretical quantiles of order 0.025 and 0.975 of a standard Normal distribution. The dark grey area corresponds to the 95% of most central values of the simulated standardized estimates.

Along with the insertion of the between-study variance component τ^2, the heterogeneity among studies can be explained through covariates that summarize study characteristics via the meta-regression model[30]

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}_i + \varepsilon_i, \quad \boldsymbol{\beta}_i = \boldsymbol{\beta} + \eta_i, \tag{5}$$

where $\boldsymbol{x}_i$ denotes a vector of p covariates available at the aggregated meta-analysis level for each study and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{\mathrm{P}})^\top$ denotes the vector of corresponding coefficients. The DerSimonian and Laird procedure can be straightforwardly extended to inference in meta-regression.

Example 2: Meat consumption data. High levels of meat consumption are suspected to be related to the increase of chronic diseases. Keeping with a substantial literature on this topic, Larsson and Orsini[31] investigate the association between unprocessed red meat and processed meat consumption with the relative risk of all-cause mortality. The authors perform a meta-analysis of prospective studies by distinguishing unprocessed red meat consumption (eight studies) and processed meat consumption (eight studies). The forest plot of the data is reported in Figure 3, while details about the studies inserted in the meta-analysis are summarized in Table 5. We consider the meta-regression model (5), with a binary covariate x_i indicating the type of meat consumption (unprocessed red *vs* processed). The inferential interest is on the significance of the coefficient β_1 associated to the type of meat consumption. The DerSimonian and Laird estimate of the coefficient associated to the meat type is 0.10 with standard error 0.051. The meta-regression based on the DerSimonian and Laird approach concludes for a doubtful association between the type of meat consumption with the risk of all-cause mortality, with an associated P-value equal to 0.054 and 95% confidence interval (0.00, 0.20), see Figure 3.

Similarly to the previous example, a simulation exercise with 1 000 replicates shows that the DerSimonian and Laird procedure overestimates the probability of Type I error, which is equal to 9.1% instead of the nominal 5%. □

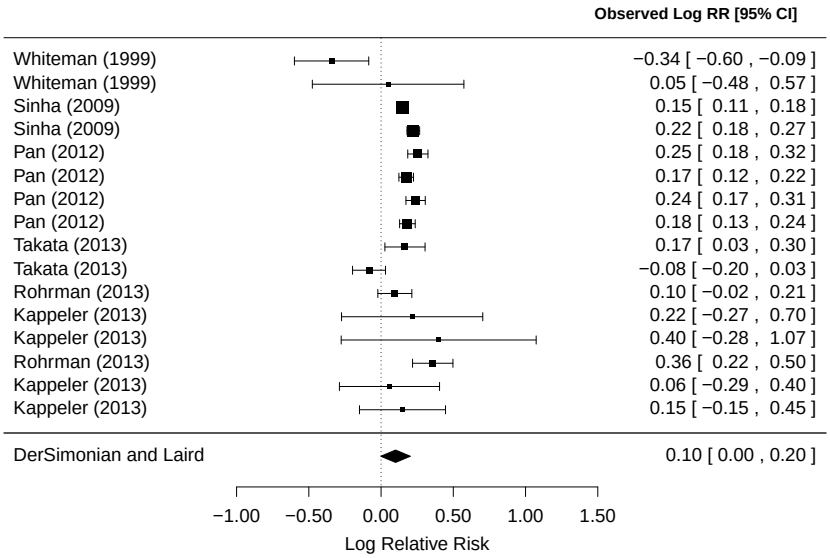

Figure 3: Forest plot of meat consumption data[31]. Outcomes from the meta-regression studies are reported in terms of the logarithm of the relative risk (Log RR), together with the associated 95% confidence interval in square brackets (95% CI). The DerSimonian and Laird estimate of the parameter associated to the type of meat consumption is reported together with the 95% confidence interval in square brackets.

3 Advanced methods for meta-analysis

Many alternatives to the DerSimonian and Laird approach have been proposed in the literature to account for the uncertainty in estimating the study heterogeneity. Proposals range from modifications of the method of moments to sophisticated solutions based on the likelihood principle, until nonparametric approaches aimed at avoiding distributional assumptions.

3.1 The Hartung and Knapp method

Hartung and Knapp[38–40] develop a modification of the DerSimonian and Laird procedure to handle small number of studies n. The procedure exploits (i) a small-sample adjustment of $\text{var}(\hat{\beta}_{\text{DL}})$ and (ii) the t distribution in place of the Normal distribution

Study	Deaths	Cohort	Type of meat
Whiteman (1999)[32]	598	10,522	Both
Sinha (2009)[33]	47,796	322,263	Processed
Sinha (2009)[33]	23,276	223,390	Processed
Pan (2012)[34]	8,926	37,698	Unprocessed
Pan (2012)[34]	15,000	83,644	Unprocessed
Takata (2013)[35]	2,733	61,128	Unprocessed
Takata (2013)[35]	4,210	73,162	Unprocessed
Rohrman (2013)[36]	26,344	448,568	Both
Kappeler (2013)[37]	1,908	8,239	Both
Kappeler (2013)[37]	1,775	9,372	Both

Table 2: Characteristics of the studies included in the meat consumption data[31].

for the standardized estimator of β. The small-sample adjusted estimator of $\text{var}(\hat{\beta}_{\text{DL}})$ is the weighted average

$$\text{var}(\hat{\beta}_{\text{DL}}) = \frac{1}{n-1} \frac{\sum_{i=1}^{n} w_i (Y_i - \hat{\beta}_{\text{DL}})^2}{\sum_{i=1}^{n} w_i}.$$

Inference on β is based on the t statistic

$$t = \frac{\hat{\beta}_{\text{DL}} - \beta}{\sqrt{\text{var}(\hat{\beta}_{\text{DL}})}}, \tag{6}$$

which follows a t distribution with $n-1$ degrees of freedom. The extension to meta-regression is investigated in Knapp and Hartung[40]. Commonly, the Hartung and Knapp method employs the DerSimonian and Laird estimator of τ^2 [38–40], although other choices are possible, e.g., empirical Bayes estimation. The Hartung and Knapp method supplies confidence intervals for β wider than those from the DerSimonian and Laird procedure with a substantial improvement in coverage accuracy. A similar methodology is suggested by Sidik and Jonkman[41]. In previous years, Berkey, Hoaglin, Mosteller et al.[42] employed an empirical Bayes estimator of τ^2 and a t distribution with $n-4$ degrees of freedom for inferential purposes. Despite its effectiveness, the t-based approaches gave rise to discussion in literature, since there have been concerns about the substitution of the variance components with sample estimates[43]. Moreover, Higgins and Thompson[44]

argue that t tests can be conservative in case of small n.

3.2 Likelihood analysis

Likelihood methods are considered a simple and effective approach to account for the uncertainty in estimating the heterogeneity by several authors[3,5,45]. Thereafter, we will concentrate on meta-analysis, although extensions to meta-regression with any number of covariates are straightforward. The log-likelihood function for $\boldsymbol{\theta} = (\beta, \tau^2)^\top$ is

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\log(\tau^2 + \sigma_i^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - \beta)^2}{\tau^2 + \sigma_i^2}. \tag{7}$$

As for the DerSimonian and Laird approach, the maximum likelihood estimator (MLE) of β is a weighted average of Y_i with weights proportional to $(\hat{\tau}^2_{\text{MLE}} + \sigma_i^2)^{-1}$, where $\hat{\tau}^2_{\text{MLE}}$ is the maximum likelihood estimate of τ^2. Since $\hat{\tau}^2_{\text{MLE}}$ depends itself on β, then the parameters need to be jointly estimated by maximization of the likelihood. Hardy and Thompson[45] and Brockwell and Gordon[3] describe a recursive algorithm for computation of the maximum likelihood estimate. Let $\hat{\boldsymbol{\theta}}_{\text{MLE}} = (\hat{\beta}_{\text{MLE}}, \hat{\tau}^2_{\text{MLE}})^\top$ be the whole parameter vector. The variance of $\hat{\beta}_{\text{MLE}}$ is estimated as $\widehat{\text{var}}(\hat{\beta}_{\text{MLE}}) = \mathbf{I}^{-1}_{\beta\beta}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$, where $\mathbf{I}_{\beta\beta}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$ denotes the element of the expected information matrix $\mathbf{I}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$ corresponding to β. The simplest likelihood approach for inference on β is based on Wald-type statistic $\widehat{\text{var}}(\hat{\beta}_{\text{MLE}})^{-1/2}(\hat{\beta}_{\text{MLE}} - \beta)$, which has a standard Normal distribution asymptotically on n. A drawback of the maximum likelihood method is related to the estimation of τ^2, since $\hat{\tau}^2_{\text{MLE}}$ may suffer from considerable downward bias[46,47], thus affecting the accuracy of the Wald statistic.

Signed profile log-likelihood ratio and Skovgaard's statistic. The Wald test may be inconvenient because *(i)* it is not invariant with respect to model reparametrization and *(ii)* confidence intervals for β are forced to be symmetric. Both the limitations can be overcome by relying on the signed profile log-likelihood ratio. Let $\hat{\tau}^2_\beta$ denote the constrained maximum likelihood estimate of τ^2 for a fixed value of β. Inference on β accounting for the variability in estimation of τ^2 can be based on the signed profile

log-likelihood ratio

$$r_{\mathrm{P}}(\beta) = \mathrm{sign}(\hat{\beta}_{\mathrm{MLE}} - \beta)\sqrt{2\{\ell_{\mathrm{P}}(\hat{\beta}_{\mathrm{MLE}}) - \ell_{\mathrm{P}}(\beta)\}},$$

where $\ell_{\mathrm{P}}(\beta) = \ell(\beta, \hat{\tau}^2_\beta)$ is the profile log-likelihood function for β. The signed profile log-likelihood ratio has an approximate standard Normal distribution up to an error of order $O(n^{-1/2})$ under mild regularity conditions, see Section 4.4 of Severini[48]. A $100(1-\alpha)\%$ confidence interval for β is given by the values satisfying $z_{\alpha/2} < r_{\mathrm{P}}(\beta) < z_{1-\alpha/2}$, with z_α being the αth quantile of a standard Normal variable. Hypothesis testing on β at significance level $100\alpha\%$ is performed by comparing $r_{\mathrm{P}}(\beta)$ to $z_{\alpha/2}$ and $z_{1-\alpha/2}$.

Inference based on r_{P} can be questionable for small sample size, since the asymptotic Normal distribution can be inaccurate. Guolo[6] proposes to refine the results by relying on a modification of r_{P} given by the Skovgaard's statistic[49], which improves convergence to the standard Normal distribution reaching a second-order accuracy $O(n^{-1})$. The Skovgaard's statistic is defined as

$$r^*_{\mathrm{P}}(\beta) = r_{\mathrm{P}}(\beta) + \frac{1}{r_{\mathrm{P}}(\beta)} \log \frac{u(\beta)}{r_{\mathrm{P}}(\beta)},$$

where term $u(\beta)$ involves various likelihood quantities, see Guolo[6] for details. Despite the apparent complexity, r^*_{P} assumes a computationally attractive form for meta-analysis and meta-regression. Simulation results highlight a substantial improvement of inferential conclusions based on r^*_{P} with respect to r_{P}. Empirical coverages of confidence intervals as well as empirical rejection rates are closer to nominal levels in small to moderate sample sizes. The difference in computational effort between r^*_{P} and r_{P} is irrelevant.

Bartlett's correction. Huizenga, Visser and Dolan[50] focus on the log-likelihood ratio statistic $W(\beta) = r^2_{\mathrm{P}}(\beta)$, which is asymptotically distributed as χ^2_1. A $100(1-\alpha)\%$ confidence interval for β is given by the values satisfying $W(\beta) < \chi^2_{1;1-\alpha}$ with $\chi^2_{1;1-\alpha}$ being the $(1-\alpha)$-th quantile of χ^2_1. A hypothesis on β can be tested at significance level $100\alpha\%$ by comparing W to $\chi^2_{1;1-\alpha}$. As for r_{P}, inference based on W can be affected for

small n. Huizenga, Visser and Dolan[50] evaluate the efficacy of the Bartlett's correction to ameliorate the χ^2_1 approximation. The Bartlett's correction replaces the test statistic W with $(1+A)^{-1}W$, where A is a function of the within- and the between-study variances. Simulation studies[50] using the raw mean difference as outcome show that the Bartlett's correction provides a Type I error rate close to the nominal level and a satisfactory power. The main disadvantage of the Bartlett's correction is related to the evaluation of A, which is somehow involved, especially in case of meta-regression.

Restricted maximum likelihood. The negative bias of the maximum likelihood estimate of τ^2 is a consequence of failing to properly account for the loss of degrees of freedom due to the estimation of the fixed-effects[51]. Restricted maximum likelihood is a popular method to correct for the degrees of freedom lost in the estimation of variance components[46,47]. The restricted maximum likelihood estimate of τ^2 is the maximizer of the marginal log-likelihood of the residuals $r_i = y_i - \hat{\beta}_{\text{MLE}}$,

$$\ell_{\text{REML}}(\tau^2) = -\frac{1}{2}\sum_{i=1}^{n}\log(\hat{\sigma}_i^2+\tau^2) - \frac{1}{2}\log\sum_{i=1}^{n}\frac{1}{(\hat{\sigma}_i^2+\tau^2)} - \frac{1}{2}\sum_{i=1}^{n}\frac{r_i^2}{\hat{\sigma}_i^2+\tau^2}.$$

Inference on β proceeds with the REML estimate of τ^2 substituting τ^2_{DL} in formula (4).

3.3 Nonparametric approaches

Nonparametric approaches have been proposed with the aim of avoiding distributional assumptions on the variables involved in meta-analysis. In particular, a quoted criticism of meta-analysis is the normality assumption for the random-effects component, see, for example, Van Houwelingen, Arends and Stijnen[5], Ghidey, Lesaffre and Stijnen[52], Kontopantelis and Reeves[53], Guolo[54]. Nonparametric approaches gain some advantages in robustness with respect to model misspecification and in controlling the probability of Type I error compared to standard meta-analysis approaches. Nevertheless, advantages are typically paid in terms of computational effort and a possible loss of power of hypothesis tests.

Permutation test. The permutation test for the significance of β carries out a re-arrangement of the data under the null hypothesis of the absence of effect. For each of M data permutations, the test statistic of interest, e.g., t statistic (6), is computed. The two-sided P-value is obtained as twice the proportion of the permutation statistics exceeding the value of the statistic computed at the observed data. Follmann and Proschan[55] carry out inference in meta-analysis by permuting the signs of the observed effect sizes, since positive and negative values are equally likely under the null hypothesis. In meta-regression, Higgins and Thompson[44] consider the permutation of the rows of the design matrix. Simulation studies in a small sample scenario highlight a good performance of the permutation method in preserving the probability of Type I error compared to t-based approaches. A limitation of the permutation test is that conventional levels of statistical significance, such as the traditional 0.05, may not be reached when the number of studies is very small[56]. For example, the smallest possible P-value is 0.0625 when $n = 5$, and 0.031 when $n = 6$. Therefore, conclusions have to be properly interpreted.

Resampling. Huizenga, Visser and Dolan[50] consider a form of residual permutation for testing the significance of β. The procedure starts with the estimate computed under the null hypothesis of the absence of effect. Then, a replicate of the data is obtained by resampling the n residuals without replacement. The distribution of the test statistic under the null hypothesis is evaluated at M data replicates. The P-value is computed as the proportion of the resampled test statistics exceeding the test statistic evaluated at the observed data. Simulation results[50] in case of small τ^2 for meta-analysis on raw mean difference conclude for a satisfactory performance of the resampling method in terms of Type I error rates, although at some loss of power, especially for small n. As for the permutation test, the number of studies n in meta-analysis affects the significance level, since the number of unique resampling data sets is $n!$ and the P-values are multipliers of $1/n!$.

Method	`metafor`[56]	`metalik`[57]	`metatest`[50]
DeSimonian and Laird	✓	✓	✓
Hartung and Knapp	✓	✗	✓
Wald test	✓	✗	✓
Signed profile log-likelihood ratio	✓†	✓	✓†
Skovgaard's statistic	✗	✓	✗
Bartlett's correction	✗	✗	✓
Restricted maximum likelihood	✓	✗	✗
Permutation test	✓	✗	✗
Resampling test	✗	✗	✓

Table 3: Implementation of meta-analysis methods in `R` packages `metafor`[56], `metalik`[57], `metatest`[50]. †The signed profile log-likelihood statistic is implicity available as the square root of the log-likelihood ratio statistic.

4 `R` implementation of the approaches

The discussed approaches for meta-analysis are implemented within the `R` programming language[7] in the following packages freely available at the CRAN repository cran.r-project.org/web/packages, see Table 3. The package `metafor`[56] provides a comprehensive collection of functionalities for meta-analysis and meta-regression, including DerSimonian and Laird and Hartung and Knapp methods, the likelihood ratio statistic and the permutation test[44]. Several estimators of τ^2 are available, such as method of moments, empirical Bayes, maximum and restricted maximum likelihood. The package `metaLik`[57] is devoted to likelihood inference in meta-analysis and meta-regression. Inferential procedures are based on the profile log-likelihood function and on the Skovgaard's statistic. Results from the DerSimonian and Laird procedure are available as well. The package `metatest`[50] implements several approaches to meta-analysis and meta-regression. Methods include DerSimonian and Laird and Hartung and Knapp methods, the likelihood ratio test and the Bartlett's correction, and the resampling procedure. A detailed and constantly updated list of `R` functionalities for meta-analysis is available within the CRAN Task View webpage, at URL cran.r-project.org/web/views/MetaAnalysis.html, maintained by Michael Dewey.

5 Simulation study

The performance of the various methods is investigated through an extensive simulation study inspired by the local anaesthesia data[24]. The focus of the simulation study is the comparison of both Type I error rates and power to detect effects. Special attention is paid to the impact of the number of studies n. Computations are carried out with the `R` meta-analysis packages listed in Section 4.

Increasing values of $n \in \{5, 10, 15, 20\}$ are considered. Data for $n = 5$ are simulated from model (2) with values of the within-study variances σ_i^2 equal to those in the local anaesthesia data and the between-study heterogeneity component τ^2 set equal to the DerSimonian and Laird estimate $\hat{\tau}^2_{\text{DL}} = 1.081$. Simulations for $n = 10$ use the same setting with each σ_i^2 replicated twice. Similarly, values of σ_i^2 are recycled for $n = 15$ and $n = 20$. Investigation of power considers values of β in the set $\{0.0, \pm 0.4, \pm 0.8, \pm 1.2, \pm 1.6\}$. The number of simulation replicates is 5 000 for each scenario. Permutation and resampling tests are conducted with M=1 000 replicates. When examining the scenario with $n = 5$, the permutation test[55] is not considered for comparison, since the range of attainable significance levels is seriously restricted[56]. Results from the resampling procedure are omitted because the method performs very unsatisfactorily. Such a result partly contrasts with findings in Huizenga, Visser and Dolan[50], where the performance of the resampling test is globally adequate. In their paper, Huizenga, Visser and Dolan conclude that results may not generalize to other scenarios, in particular to effect size metrics different from the raw mean difference they focus on.

Type I error rates. Figure 4 reports the Type I error rates of the significance test on β from different approaches according to increasing sample size n, with nominal level 0.05. The poor performance of the DerSimonian and Laird approach and the Wald statistic is evident, with Type I error rates exceeding the nominal level, more seriously as n decreases. The restricted maximum likelihood only slightly reduces the overestimation of the probability of Type I error. Remaining methods perform considerably better. The Hartung and Knapp method and the Skovgaard's statistic provide a satisfactory

result irrespective of n. The Skovgaard's statistic substantially improves the signed profile log-likelihood ratio statistic r_{P}, which results in overestimation of the target level. The improvement is remarkable for small n, by this way confirming previous findings in literature[6]. The Bartlett's correction performs very satisfactorily, yielding P-values quite close to those based on the Skovgaard's statistic. The nonparametric solution provided by the permutation test, examined for all n but $n = 5$, is adequate as well.

Power. Figure 5 displays the power of the different tests as a function of β and $n \in \{5, 10, 20\}$. As expected, a superior power is held in case of small n by those methods failing to maintain the desired probability of Type I error, namely, the Wald statistic, the DerSimonian and Laird approach and the restricted maximum likelihood. Conversely, the less powerful tests are those based on the Skovgaard's statistic, the Bartlett correction and the Hartung and Knapp method. Globally, the performance of the competing approaches notably improves as n increases, with small differences among the methods for $n = 10$ and negligible differences for $n = 20$.

6 Analysis of the examples

Example 1: Local anesthesia data. Consider the meta-analysis model (2) to investigate the efficacy of local anaesthesia in reducing pain associated to hysteroscopy. The application of the DerSimonian and Laird procedure strongly supports pain reduction (P-value 0.007), see Table 4. The conclusion from the Wald test is doubtful given a P-value of 0.056. Conversely, the Hartung and Knapp method, restricted maximum likelihood, the signed profile log-likelihood ratio statistic r_{P}, the Skovgaard's statistic r_{P}^{*} and the Bartlett's correction of the log-likelihood ratio test do not support the efficacy of anaesthesia. The permutation test has not been reported given the small number of studies that does not allow to evaluate significance at 5% level.

Example 2: Meat consumption data. Consider the meta-regression model (5) to investigate the association between the risk of all-cause mortality and the consumption of

Table 4: Meta-analysis of local anaesthesia data[24]: P-values for various meta-analysis approaches.

Method	P-value
DerSimonian and Laird	0.007
Hartung and Knapp	0.170
Wald test	0.056
Signed profile log-likelihood ratio	0.096
Skovgaard's statistic	0.158
Bartlett's correction	0.144
Restricted maximum likelihood	0.087

unprocessed red or processed meat. The DerSimonian and Laird approach concludes for a dubious association between the type of meat consumption and the risk of mortality, given a P-value equal to 0.054, see Table 5. Conversely, all the alternative methods undoubtedly conclude for no difference between the risk of mortality for unprocessed red meat consumption or processed meat consumption.

Table 5: Meta-regression of meat consumption data[31]: P-values for various meta-regression approaches.

Method	P-value
DerSimonian and Laird	0.054
Hartung and Knapp	0.146
Wald test	0.082
Signed profile log-likelihood ratio	0.095
Skovgaard's statistic	0.145
Bartlett's correction	0.117
Restricted maximum likelihood	0.111
Permutation test	0.146

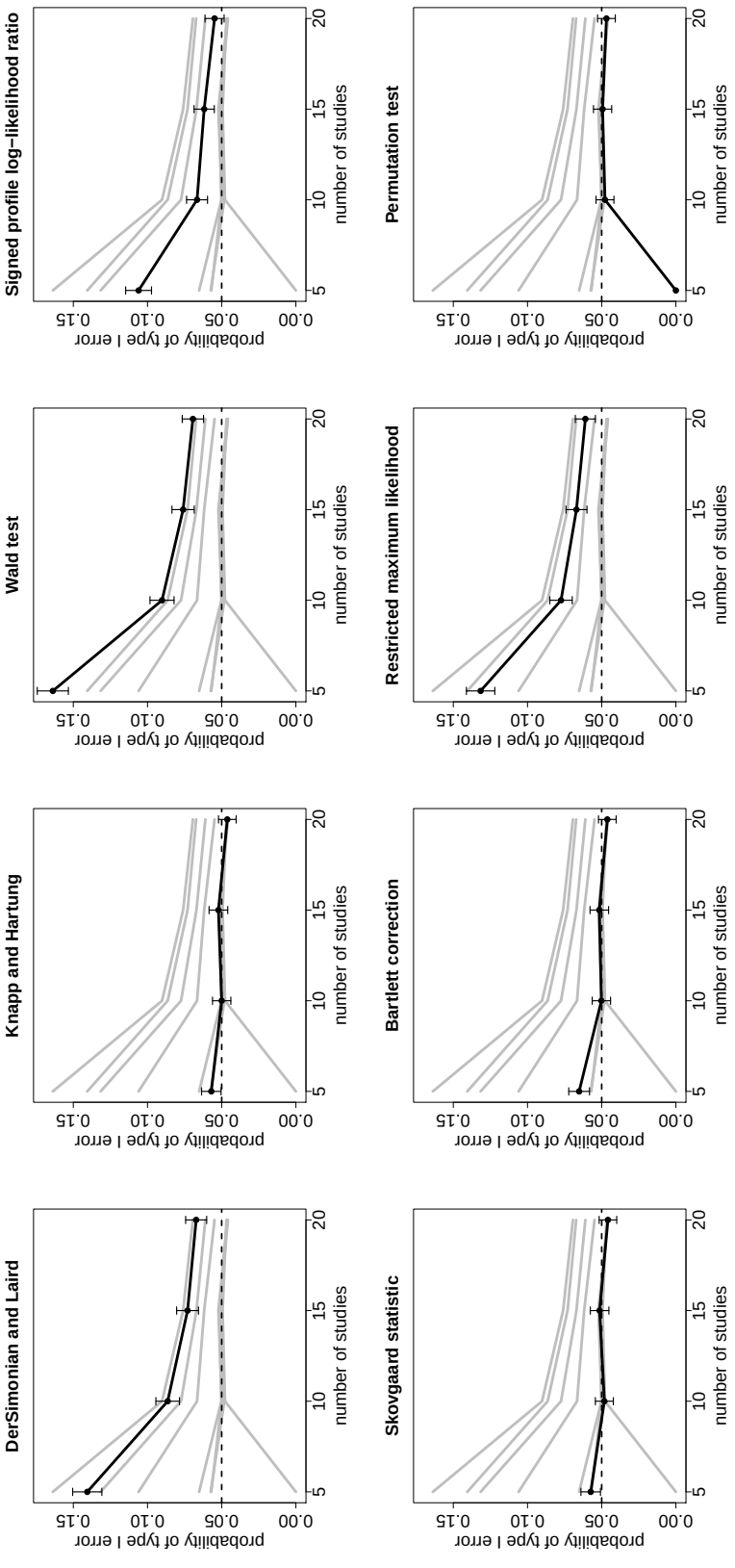

Figure 4: Simulation study. Type I error rates of various tests as a function of increasing values of the number of studies n. In each panel the horizontal dashed line corresponding to the probability of Type I error equal to 0.05 is superposed. Vertical bars identify 95% confidence intervals.

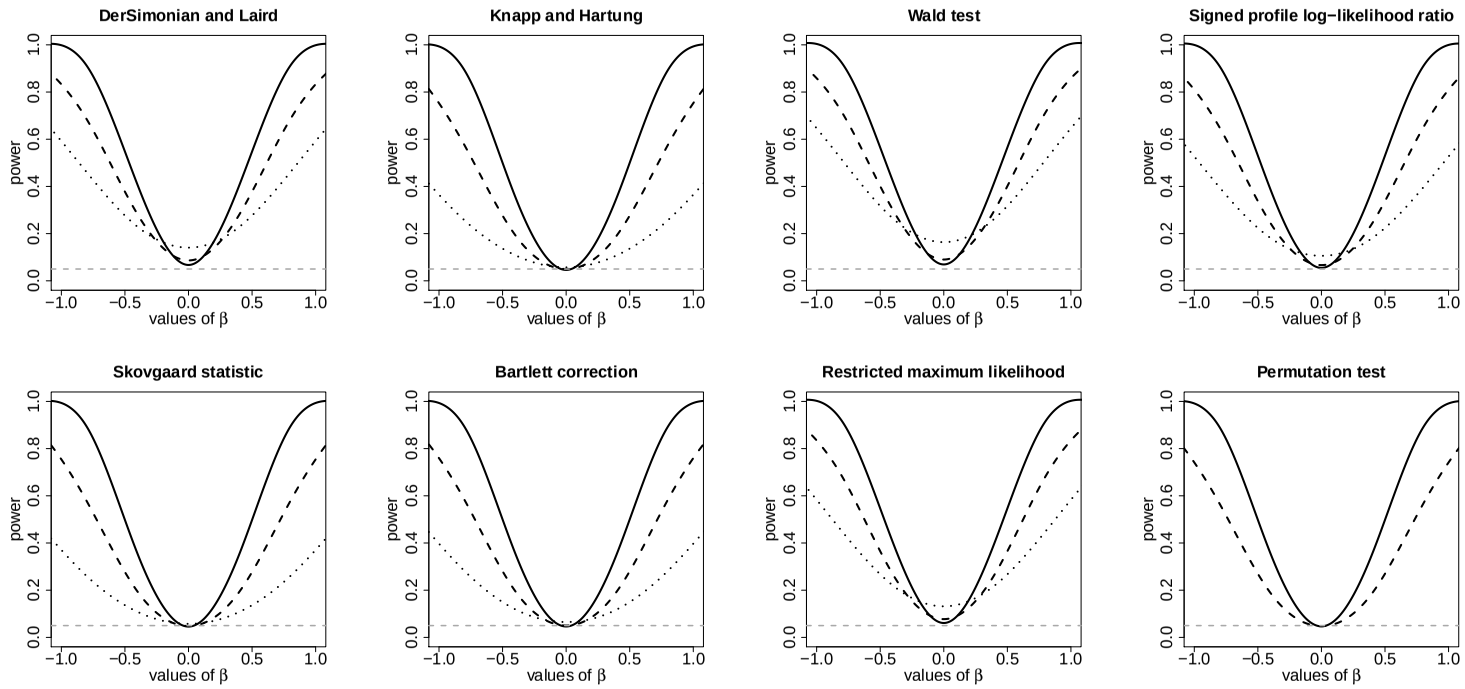

Figure 5: Simulation study. Empirical power of various tests for number of studies n equal to 5 (dotted line), 10 (dashed line), 20 (solid line). In each panel the horizontal grey dashed line corresponding to the significance level 0.05 is superposed.

7 Concluding remarks

Keeping with a substantial portion of the recent literature in meta-analysis, the inaccuracy of the traditional DerSimonian and Laird approach has been empirically shown in terms of coverage of confidence intervals and P-values. In this paper, several alternative methods have been compared with attention to their performance in case of small number of studies. The performance of the methods has been investigated in terms of both Type I error rates and empirical power for detecting effects.

Although results do not allow to obtain an overall conclusion about a superior method whatever the scenario, the study provides some interesting results. Within the likelihood methods, sophisticated solutions, such as the Skovgaard's statistic or the Bartlett's correction, are preferable to the traditional Wald approach or to the signed profile log-likelihood ratio statistic in maintaining desired levels of the probability of Type I error. The price to pay is a slightly reduced power when the number of studies is small. In case a nonparametric solution is chosen given doubts about distributional assumptions, the permutation test appears to be preferable to the resampling test in terms of Type I error rates, for different sample sizes. Drawbacks of the permutation approach involve a slightly reduced power and a substantial computational effort as the number of studies increases. When the investigator intends to maintain the simplicity of the DerSimonian and Laird approach avoiding the complexity of other procedures, the Hartung and Knapp adjustment may represent a good compromise.

Bayesian solutions have not been examined in this paper, although they can represent a viable alternative approach to meta-analysis[58]. The `R` package `metamisc`[59] implements Bayesian meta-analysis under the normality assumption for the random-effects. The current package version does not consider meta-regression.

The recommendation arisen from this paper is that a comprehensive meta-analysis requires the comparison of different methods, especially in the frequent case of a small number of studies. Such a strategy, although more elaborated than the use of a single technique, is actually not involved, given the availability of several `R` packages for performing meta-analysis.

Supplementary material

The `R` code for replication of examples is provided as supplementary material.

Conflict of interest statement

The Authors declare that there is no conflict of interest.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors thank Dr. Ioannis Kosmidis for discussion.

References

1. Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ* 2001; **322** (7278): 98–101.

2. DerSimonian R and Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.

3. Brockwell SE and Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20** (6): 825–840.

4. DerSimonian R and Kacker R. Random-effects model for meta-analysis of clinical trials: An update. *Control Clin Trials* 2007; **28**: 105–114.

5. Van Houwelingen HC, Arends LR and Stijnen T. Tutorial in biostatistics. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; **21** (4): 589–624.

6. Guolo A. Higher-order likelihood inference in meta-analysis and meta-regression. *Stat Med* 2012; **31** (4): 313–327.

7. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria; 2014. http://www.R-project.org/.

8. Schmidt FL, Oh I-S and Hayes TL. Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *Brit J Math Stat Psy* 2009; **62**: 97–128.

9. Borenstein M, Hedges LV, Higgins JPT and Rothstein HR. A basic introduction to fixed-effects and random-effects models for meta-analysis. *Research Synthesis Methods* 2010; **1**: 97–111.

10. Hunter JE and Smith FL. Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *Int J Select Assess* 2000; **8**: 275–292.

11. Schulze R. *Meta-analysis: A comparison of approaches*. Hogrefe & Huber: Toronto, 2004.

12. Higgins JPT and Green S (editors). *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

13. Hedges LV and Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486–504.

14. Hardy RJ and Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med* 1998; **17** (8): 841–856.

15. Whitehead A. *Meta-analysis of controlled clinical trials*. Wiley: Chichester, 2002.

16. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. *Br J Math Stat Psychol* 2007; **60** (1): 29–60.

17. Cochran WG. Problems arising in the analysis of a series of similar experiments. *J R Stat Soc Supplement* 1937; **4**: 102–118.

18. Hedges LV and Olkin I. *Statistical methods for meta-analysis*. San Diego, CA: Academic Press, 1985.

19. Galbraith RF. A note on graphical presentation of estimated odds ratios from several trials. *Stat Med* 1988; **7**: 889–894.

20. Julious SA and Whitehead A. Investigating the assumption of homogeneity of treatment effects in clinical studies with application to meta-analysis. *Pharm Stat* 2012; **11**: 49–56.

21. Higgins JPT and Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–1558.

22. Higgins JPT, Thompson SG, Deeks JJ and Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.

23. Roloff V, Higgins JPT and Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat Med* 2013; **32** (1): 11–24.

24. Cooper NAM, Khan KS and Clark TJ. Local anaesthesia for pain control during outpatient hysteroscopy: systematic review and meta-analysis. *BMJ* 2010; **340**: c1130.

25. Vercellini P, Colombo A, Mauro F, Oldani S, Bramante T and Crosignani PG. Paracervical anesthesia for outpatient hysteroscopy. *Fertil Steril* 1994; **62**: 1083–1085.

26. Cicinelli E, Didonna T, Schonauer LM, Stragapede S, Falco N and Pansini N. Paracervical anesthesia for hysteroscopy and endometrial biopsy in postmenopausal women. A randomized, double-blind, placebo-controlled study. *J Reprod Med* 1998; **43**: 1014–1018.

27. Lau WC, Lo WK, Tam WH and Yuen PM. Paracervical anaesthesia in outpatient hysteroscopy: a randomised double-blind placebo- controlled trial. *Br J Obstet Gynaecol* 1999; **106**: 356–359.

28. Giorda G, Scarabelli C, Franceschi S and Campagnutta E. Feasibility and pain control in outpatient hysteroscopy in postmenopausal women: a randomized trial. *Acta Obstet Gynecol Scand* 2000; **79**: 593–597.

29. Al-Sunaidi M and Tulandi T. A randomized trial comparing local intracervical and combined local and paracervical anesthesia in outpatient hysteroscopy. *J Minim Invasive Gynecol* 2007; **14**: 153–155.

30. Thompson SG and Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; **21** (11): 1559–1573.

31. Larsson SC and Orsini N. Red meat and processed meat consumption and all-cause mortality: A meta-analysis. *Am J Epidemiol* 2014; **179** (3): 282–289.

32. Whiteman D, Muir J, Jones L, Murphy M and Key T. Dietary questions as determinants of mortality: the OXCHECK experience. *Public Health Nutr* 1999; **2** (4): 477–487.

33. Sinha R, Cross AJ, Graubard BI, Leitzmann MF and Schatzkin A. Meat intake and mortality: a prospective study of over half a million people. *Arch Intern Med* 2009; **169** (6): 562–571.

34. Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Stampfer MJ, Willett WC and Hu FB. Red meat consumption and mortality: results from 2 prospective cohort studies. *Arch Intern Med* 2012; **172** (7): 555–563.

35. Takata Y, Shu XO, Gao YT, Li H, Zhang X, Gao J, Cai H, Yang G, Xiang Y-B and Zheng W. Red meat and poultry intakes and risk of total and cause-specific mortality: results from cohort studies of Chinese adults in Shanghai. *PLoS One* 2013; **8** (2): e56963.

36. Rohrmann S, Overvad K, Bueno-de-Mesquita HB, et al. Meat consumption and mortality–results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med* 2013; **11**: 63.

37. Kappeler R, Eichholzer M and Rohrmann S. Meat consumption and diet quality and mortality in NHANES III. *Eur J Clin Nutr* 2013; **67** (6): 598–606.

38. Hartung J and Knapp G. On tests of the overall treatment effect in the meta-analysis with normally distributed responses. *Stat Med* 2001; **20** (12): 1771–1782.

39. Hartung J and Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 2001; **20** (24): 3875–3889.

40. Knapp G and Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; **22** (17): 2693–2710.

41. Sidik K and Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med* 2002; **21** (21): 3153–3159.

42. Berkey CS, Hoaglin DC, Mosteller F et al. A random-effects regression model for meta-analysis. *Stat Med* 1995; **14** (4): 395–411.

43. Copas J. A simple confidence interval for meta-analysis. Sidik K, Jonkman JN. *Stat Med* 2002; **21** (21): 3153–9. [letter] *Stat Med* 2003; **22** (16): 2667–2668.

44. Higgins JPT and Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004; **23** (11): 1663–1682.

45. Hardy RJ and Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med* 1996; **15** (6): 619–629.

46. Patterson HD and Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika* 1971; **58** (3): 545–554.

47. Patterson HD and Thompson R. Maximum likelihood estimation of components of variance. In: *Proceedings of the 8th International Biometrics Conference*, Biometric Society, Washington, DC, 1974, pp. 197–207.

48. Severini TA. *Likelihood methods in statistics*. Oxford: Oxford University Press, 2000.

49. Skovgaard IM. An explicit large-deviation approximation to one-parameter tests. *Bernoulli* 1996; **2** (2): 145–165.

50. 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): 1–19.

51. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J Educ Behav Stat* 2005; **30**: 261–293.

52. Ghidey W, Lesaffre E and Stijnen T. Semi-parametric modelling of the distribution of the baseline risk in meta-analysis. *Stat Med* 2007; **26**: 5434–5444.

53. Kontopantelis E and Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Stat Methods Med Res* 2010; **21** (4): 409–426.

54. Guolo A. Flexibly modelling the baseline risk in meta-analysis. *Stat Med* 2012; **32** (1): 40–50.

55. Follmann DA and Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999; **55** (3): 732–737.

56. Viechtbauer W. Conducting meta-analyses in R with the `metafor` package. *J Stat Softw* 2010; **36** (3): 1–48.

57. Guolo A and Varin C. The R package `metaLik` for likelihood inference in meta-analysis. *J Stat Softw* 2012; **50** (7): 1–14.

58. Higgins JPT, Thompson SG and Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Series A* 2009; **172** (1): 137–159.

59. Debray T. metamisc: Diagnostic and prognostic meta analysis (metamisc). R package version 0.1.1, 2013. http://CRAN.R-project.org/package=metamisc.

Appendix

The various approaches to meta-analysis compared in this paper are implemented within the different `R` packages listed in Section 4. In order to facilitate the simultaneous use of the several meta-analysis methods advocated in this paper, a single `R` function called `metamany` with a user-friendly syntax is provided in the supplementary material. Function `metamany` requires previous installation of `R` packages `metafor`[56], `metaLik`[57] and `metatest`[50]

```
R> install.packages(metafor)
R> install.packages(metaLik)
R> install.packages(metatest)
```

Once installed the three packages above, function `metamany` can be loaded

```
R> source("metamany.R")
```

The arguments of function `metamany` are

```
metamany(y, sigma2, X = NULL, param = NULL)
```

where `y` and `sigma2` are the vectors of estimated outcomes and within-study variances, respectively. Optional input `X` allows to specify a $n \times p$ matrix of study-specific covariates (intercept excluded) for meta-regression. Optional input `param` allows to specify which parameter should be tested in meta-regression. If `param` is left unspecified, then the parameter corresponding to the last column of `X` is tested.

Local anesthesia data are available through data frame `cooper`:

```
R> cooper
      y  sigma2
1  0.00 0.03959
2 -1.71 0.07732
3 -0.19 0.02265
4 -0.58 0.01760
5 -4.27 0.16041
```

Since nonparametric methods use resampling, thereafter the random seed is fixed to allow the reproducibility of the results:

```
R> set.seed(0207)
R> metamany(y = cooper$y, sigma2 = cooper$sigma2)


Estimates:
                              Estimate  Std.Err.  Heterogeneity
DerSimonian and Laird          -1.283     1.081     0.478
Maximum likelihood             -1.324     2.927     0.773
Restricted maximum likelihood  -1.317     2.306     0.688
```

```
P-values:

                                    P-value
DerSimonian and Laird               0.00727
Hartung and Knapp                   0.17044
Wald test                           0.05580
Signed profile log-likelihood ratio 0.09604
Skovgaard statistic                 0.15846
Bartlett correction                 0.14441
Restricted maximum likelihood       0.08695
Permutation test                    0.12500
```

Warning: Given the number of studies, the P-value of the permutation test does not allow to evaluate significance at 5% level.

Meat consumption data are available through data frame `larsson`:

```
R> larsson
         y   sigma2 type
1  -0.3425 0.017224    a
2   0.2546 0.001271    a
3   0.1740 0.000663    a
4   0.1655 0.005027    a
...
```

where variable `type` distinguishes between unprocessed red meat (`type a`) and processed meat (`type b`) consumption.

Meta-regression of meat consumption data with the random seed fixed:

```
R> set.seed(0207)
R> metamany(y = larsson$y, sigma2 = larsson$sigma2,  X = larsson$type)

Estimates:
                                Estimate  Std.Err.  Heterogeneity
```

DerSimonian and Laird	0.10044	0.00567	0.05218
Maximum likelihood	0.10975	0.01184	0.06891
Restricted maximum likelihood	0.10639	0.00850	0.06117

P-values:

	P-value
DerSimonian and Laird	0.0543
Knapp and Hartung	0.1459
Wald test	0.0820
Signed profile log-likelihood ratio	0.0946
Skovgaard statistic	0.1454
Bartlett correction	0.1170
Restricted maximum likelihood	0.1113
Permutation test	0.1460