Available online at www.sciencedirect.com
ScienceDirect
Contemporary Clinical Trials 28 (2007) 105–114
Contemporary Clinical Trials
www.elsevier.com/locate/conclintrial

Random-effects model for meta-analysis of clinical trials: An update

Rebecca DerSimonian [a,*], Raghu Kacker [b]

[a] *Biostatistics Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA*
[b] *Mathematical and Computational Sciences Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA*

Received 7 April 2005; accepted 12 April 2006

Abstract

The random-effects model is often used for meta-analysis of clinical studies. The method explicitly accounts for the heterogeneity of studies through a statistical parameter representing the inter-study variation. We discuss several iterative and non-iterative alternative methods for estimating the inter-study variance and hence the overall population treatment effect. We show that the leading methods for estimating the inter-study variance are special cases of a general method-of-moments estimate of the inter-study variance. The general method suggests two new two-step methods. The iterative estimate is statistically optimal and it can be easily calculated on a spreadsheet program, such as Microsoft Excel, available on the desktop of most researchers. The two-step methods approximate the optimal iterative method better than the earlier one-step non-iterative methods.

Keywords: Clinical trials; Meta-analysis; Random effects model; Restricted maximum likelihood; Variance components

1. Introduction

Meta-analysis is a statistical technique for combining estimated treatment effects from independent comparable clinical trials (studies). Such analyses have become increasingly popular in medical research where information about treatment efficacy is available from a number of clinical trials with inconclusive or inconsistent results.

A major difficulty in integrating the findings from various studies stems from the sometimes diverse nature of the studies being combined. The studies may differ, for example, in terms of patient characteristics or methods employed. To account for such inter-study differences, DerSimonian and Laird [1] proposed a simple random effects model which allows for treatment effects to vary across studies and uses a simple non-iterative method to estimate the inter-study treatment effect variance. Because it incorporates inter-study differences into the analysis of overall treatment efficacy, and because of its simplicity, the method [1] continues to be widely used. Nevertheless, indiscriminate or inappropriate use of any approach to meta-analysis of clinical trials can lead to misleading inferences about treatment effects [2], and

* Corresponding author. Tel.: +1 301 435 7181; fax: +1 301 480 0912.
E-mail address: DerSimonian@nih.gov (R. DerSimonian).

1551-7144/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2006.04.004

the need for careful consideration of methods in drawing statistical inferences from comparable but heterogeneous studies remains critical.

In this paper, we first review the random-effects model for meta-analysis of clinical trials and introduce a general method-of-moments estimate for the inter-study variance which includes several existing estimates as special cases. In addition to the non-iterative method proposed by DerSimonian and Laird [1], an iterative estimate of the inter-study variance based on a random effects model was proposed by Paule and Mandel [3] for inter-laboratory studies. This estimate was subsequently shown to be statistically optimal [4] and can be easily calculated on a spreadsheet program. Another non-iterative estimate of the inter-study variance component based on a random-effects model was proposed by Cochran [5]. In contrast to the non-iterative DerSimonian and Laird as well as the iterative Paule and Mandel estimates, the estimate based on Cochran's ANOVA assumes that each study provides equal information and is of equal sample size.

We show that the inter-study variance estimates based on the methods of Cochran, DerSimonian and Laird, and Paule and Mandel are all special cases of a general method-of-moments estimate for the inter-study variance with slightly different weights assigned to the studies. The general method-of-moments estimate suggests two-step alternatives to the one-step non-iterative procedures based on Cochran's ANOVA and the DerSimonian and Laird methods. We illustrate and compare the estimates from the five methods in several examples, and based on the empirical evidence, suggest improvements to the commonly used one-step non-iterative random-effects model estimates.

2. Methods

We consider the problem of combining estimated treatment effects from a series of k comparative clinical studies, where the data from each study consist of the number of patients in treatment and control groups, n_{Ti} and n_{Ci}, and the proportion of patients with some event in each of the two groups, r_{Ti} and r_{Ci}. A random effects model for meta-analysis stipulates that the observed treatment effect, y_i, from the i-th clinical study is made up of two additive components: the true treatment effect for the study, θ_i, and the sampling error, e_i. That is, $y_i=\theta_i+e_i$ for $i=1,\ldots,k$. The variance of e_i, σ_i^2, is the sampling variance reflecting within-study variance and the sample size of the study. The sampling variance, σ_i^2, is usually unknown and is estimated from the data of the i-th observed study. For instance, when the observed effect in the i-th study is a difference in proportions, $r_{Ti}-r_{Ci}$, the sampling variance can be estimated [1] by

$$s_i^2 = r_{Ti}(1-r_{Ti})/n_{Ti} + r_{Ci}(1-r_{Ci})/n_{Ci.}$$

In addition to the sampling error associated with each study, the random effects model assumes the true treatment effect in each trial will be influenced by several factors, including patient characteristics as well as design and execution of the study. The model explicitly accounts for this possible heterogeneity in the true treatment effects and stipulates that $\theta_i=\mu+\delta_i$, where θ_i is the true treatment effect in the i-th study, μ is the overall treatment effect for a population of possible treatment evaluations, and $\delta_i=\theta_i-\mu$ is the deviation of the i-th study's effect from the overall effect μ. The variance of δ_i, $\tau^2\geq 0$, is the inter-study variance and represents both the degree to which true treatment effects vary across experiments as well as the degree to which individual studies give biased assessments of treatment effects. The special case $\tau^2=0$ represents lack of heterogeneity among the true treatment effects; i.e., the true treatment effects θ_i are all equal and the common value is μ.

With this formulation, the model assumes that the observed treatment effects, $y_1,\ldots, y_k$, are realizations of independent random variables from a distribution with overall value μ and variances $\tau^2+\sigma_1^2,\ldots, \tau^2+\sigma_k^2$, respectively, where $\sigma_1^2>0,\ldots, \sigma_k^2>0$ and $\tau^2\geq 0$. The variances reflect the two components of variance assigned to each observed effect: an inter-study variance τ^2 which reflects treatment effects heterogeneity and an intra-study variance σ_i^2 (or its approximation s_i^2) which reflects within-study sampling variance.

2.1. Estimation of the overall population treatment effect μ

Given the observed effects, $y_1,\ldots, y_k$, and the sampling variances, $\sigma_1^2,\ldots, \sigma_k^2$, the first step in meta-analysis based on a random effects model is to calculate an estimate for the inter-study variance τ^2 and then estimate the overall population treatment effect μ and its standard error.

If $\sigma_1^2,\ldots, \sigma_k^2$ and τ^2 were known, a weighted estimator of μ would be $\mu_W=\Sigma_i W_i y_i/\Sigma_i W_i$, where $W_i=1/(\tau^2+\sigma_i^2)$, and its standard error would be s.e.$(\mu_W)=1/(\Sigma_i W_i)^{1/2}$. In practice, the variances $\sigma_1^2,\ldots, \sigma_k^2$, and τ^2 are usually unknown and

are estimated from the data. Suppose $s_1^2,\ldots, s_k^2$ and t^2 are the estimates of $\sigma_1^2,\ldots, \sigma_k^2$ and τ^2, respectively. By substituting the estimated variances for $\sigma_1^2,\ldots, \sigma_k^2$, and τ^2 in μ_W we get the following estimate for μ

$$m_W = \sum_i w_i y_i / \sum_i w_i, \tag{1}$$

where

$$w_i = 1/(t^2 + s_i^2) \tag{2}$$

for $i=1,\ldots, k$ and an approximate standard error for m_w

$$\text{s.e.}(m_W) = 1/\left(\sum_i w_i\right)^{1/2}. \tag{3}$$

The expression (3) for s.e.(m_w) is a conditional standard error of m_w under the assumption that the estimates $s_1^2,\ldots, s_k^2$ and t^2 are equal to the true variances $\sigma_1^2,\ldots, \sigma_k^2$ and τ^2, respectively. It is difficult to determine an expression for the unconditional (true) standard error of m_w involving the uncertainty that arises from the use of estimates $s_1^2,\ldots, s_k^2$ and t^2 for $\sigma_1^2,\ldots, \sigma_k^2$, and τ^2, respectively. The expression (3) is an underestimate of the true standard error of m_w [6]. General inferential problems due to replacing $\sigma_1^2,\ldots, \sigma_k^2$ by their estimates $s_1^2,\ldots, s_k^2$ in the context of a random-effect model are also discussed by Bohning et al. [7].

2.2. Estimation of the inter-study variance τ^2

In addition to the sampling variance estimates, $s_1^2,\ldots, s_k^2$, the expressions (1) and (3) require an estimate t^2 for τ^2. In this section, we describe several methods for estimating τ^2 that yield slightly different results for the overall population treatment effect estimate and its standard error, m_w and s.e.(m_w). These methods include non-iterative estimates proposed by Cochran [5] and DerSimonian and Laird [1] and an iterative estimate proposed by Paule and Mandel [3] for inter-laboratory studies. This section introduces an identity that shows that all three estimates are special cases of a general method-of-moments estimate for τ^2 proposed by Kacker [8] for inter-laboratory studies and proposes two new estimates.

Suppose $y_W = \Sigma_i a_i y_i / \Sigma_i a_i$ where $a_1,\ldots, a_k$ are any positive constants. Kacker [8] verified that the expected value of the expression $\Sigma_i a_i (y_i - y_W)^2$ is

$$E\left[\sum_i a_i (y_i - y_W)^2\right] = \sum_i a_i(\tau^2 + \sigma_i^2) - \sum_i a_i^2(\tau^2 + \sigma_i^2)/\sum_i a_i. \tag{4}$$

which can be expressed as

$$E\left[\sum_i a_i (y_i - y_W)^2\right] = \tau^2\left[\sum_i a_i - \sum_i a_i^2/\sum_i a_i\right] + \left[\sum_i a_i\sigma_i^2 - \sum_i a_i^2\sigma_i^2/\sum_i a_i\right]. \tag{5}$$

2.2.1. General method-of-moments estimate for τ^2

By equating the expression $\Sigma_i a_i (y_i - y_W)^2$ to its expected value given by Eq. (5), solving for τ^2, and then substituting $s_1^2,\ldots, s_k^2$ for $\sigma_1^2,\ldots, \sigma_k^2$, we have the following general method-of-moments estimate t^2(MM) for τ^2

$$t^2(\text{MM}) = \frac{\left[\sum_i a_i (y_i - y_W)^2\right] - \left[\sum_i a_i s_i^2 - \sum_i a_i^2 s_i^2/\sum_i a_i\right]}{\left[\sum_i a_i - \sum_i a_i^2/\sum_i a_i\right]} \tag{6}$$

Since $\tau^2 \geq 0$, the estimate τ^2(MM) is set to zero when its computed value turns out to be negative [8].

In Eq. (6), $a_1,\ldots, a_k$ are any positive values reflecting weights assigned to the k studies. Each set of values for the weights yields an alternative estimate for τ^2. Table 1 summarizes five alternate sets of weights which yield special cases of the general method of moments estimate of τ^2 described in Eq. (6). These special cases include Cochran's [5] analysis-of-variance (ANOVA) non-iterative estimate, t^2(CA), DerSimonian and Laird's [1] non-iterative estimate,

Table 1
Various estimates for τ^2 as special cases of the general method of moments estimate

Weights a_i	Estimate t^2	Author(s)	Reference
Any positive value	t^2(MM) Eq. (6)	Kacker [8]	Metrologia (2004)
$a_i=1/k$	t^2(CA) Eq. (7)	Cochran [5]	Biometrics (1954)
$a_i=1/s_i^2$	t^2(DL) Eq. (8)	DerSimonian and Laird [1]	Cont Clin Trials (1986)
$a_i=1/(t^2(\mathrm{PM})+s_i^2)$	t^2(PM)[a] Eqs. (9)–(10)	Paule and Mandel [3]	J Res NBS (1982)
$a_i=1/(t^2(\mathrm{CA})+s_i^2)$	t^2(CA2) Eq. (11)	DerSimonian and Kacker	This paper
$a_i=1/(t^2(\mathrm{DL})+s_i^2)$	t^2(DL2) Eq. (12)	DerSimonian and Kacker	This paper

[a] Requires iteration.

t^2(DL), Paule and Mandel's [3] iterative estimate, t^2(PM), two-step estimate starting with Cochran's estimate of τ^2, t^2(CA2), and two-step estimate starting with the DerSimonian and Laird estimate of τ^2, t^2(DL2). The following sections describe the five alternative sets of weights and the corresponding τ^2 estimates in more detail.

2.2.2. Cochran ANOVA estimate for τ^2

With $a_i=1/k$ for $i=1,\ldots, k$, Eq. (6) yields the Cochran ANOVA estimate, t^2(CA), for τ^2, where

$$t^2(\mathrm{CA}) = \max\left\{0, \frac{1}{k-1}\sum\nolimits_i (y_i - y_\mathrm{A})^2 - \frac{1}{k}\sum\nolimits_i s_i^2\right\}, \tag{7}$$

and y_A is the arithmetic mean $(1/k)\Sigma_i y_i$. Substituting t^2(CA) for t^2 in Eq. (2) yields the corresponding Cochran estimate, m_w(CA), for μ and its approximate standard error, s.e.(m_w(CA)), as defined by Eq. (3).

2.2.3. DerSimonian and Laird estimate for τ^2

With $a_i=1/s_i^2$ for $i=1,\ldots, k$, Eq. (6) yields the DerSimonian and Laird estimate, t^2(DL), for τ^2, where

$$t^2(\mathrm{DL}) = \max\left\{0, \frac{\left[\sum_i w_{i0}(y_i - y_\mathrm{w}(0))^2\right] - (k-1)}{\left[\sum_i w_{i0} - \sum_i w_{i0}^2 / \sum_i w_{i0}\right]}\right\}, \tag{8}$$

$y_\mathrm{W}(0)=\Sigma_i w_{i0} y_i / \Sigma_i w_{i0}$, and $w_{i0}=1/s_i^2$. Substituting t^2(DL) for t^2 in Eq. (2) yields the corresponding DerSimonian and Laird estimate, m_w(DL), for μ and its approximate standard error, s.e.(m_w(DL)), as defined by Eq. (3). The expression $\Sigma_i w_{i0}(y_i - y_\mathrm{W}(0))^2$ is referred to as a Q-statistic in meta-analysis literature [1].

2.2.4. Paule and Mandel estimate for τ^2

With $a_i=1/(\tau^2+\sigma_i^2)$ for $i=1,\ldots, k$, Eq. (4) reduces to

$$E\left[\sum\nolimits_i a_i (y_i - y_\mathrm{W})^2\right] = k-1, \tag{9}$$

where $y_\mathrm{W}=\Sigma_i a_i y_i / \Sigma_i a_i$, and $a_i=1/(\tau^2+\sigma_i^2)$. By equating the expression $\Sigma_i a_i (y_i - y_\mathrm{W})^2$, where $a_i=1/(\tau^2+\sigma_i^2)$, to its expected value $k-1$, and then substituting $s_1^2,\ldots, s_k^2$ for $\sigma_1^2,\ldots, \sigma_k^2$, we get the Paule and Mandel estimating equation

$$F(\tau^2) = \sum\nolimits_i W_i (y_i - y_\mathrm{W}(\tau^2))^2 - (k-1) = 0, \tag{10}$$

where $y_\mathrm{W}(\tau^2)=\Sigma_i W_i y_i / \Sigma_i W_i$ and $W_i=1/(\tau^2+s_i^2)$. The solution, t^2(PM), of the estimating equation $F(\tau^2)=0$ is the Paule and Mandel estimate for τ^2. The estimating equation $F(\tau^2)=0$ has a unique solution, which can be determined through a simple numerical iteration starting with $\tau^2=0$. An algorithm for numerical iteration on a spreadsheet program such as Microsoft Excel is described in the Appendix. When $F(\tau^2)$ is negative for all $\tau^2 \geq 0$, the equation $F(\tau^2)=0$ has no non-negative solution; in that case the estimate t^2(PM) is set to zero. Substituting t^2(PM) for t^2 in Eq. (2) yields the Paule and Mandel estimate, m_w(PM), for μ and its approximate standard error, s.e.(m_w(PM)), as defined by Eq. (3).

The general method of moments estimate t^2(MM) for τ^2 suggests two new two-step estimates for τ^2.

2.2.5. Two-step estimate of μ starting with the Cochran estimate for τ^2

If we substitute $a_i = 1/(t^2(\mathrm{CA}) + s_i^2)$ for $i = 1,\ldots, k$ in Eq. (6), we get the two-step estimate, t^2(CA2), where

$$t^2(\mathrm{CA2}) = \max\left\{0, \frac{\left[\sum_i w_{i\mathrm{C}}(y_i - m_{\mathrm{w}}(\mathrm{CA}))^2\right] - \left[\sum_i w_{i\mathrm{C}} s_i^2 - \sum_i w_{i\mathrm{C}}^2 s_i^2 / \sum_i w_{i\mathrm{C}}\right]}{\left[\sum_i w_{i\mathrm{C}} - \sum_i w_{i\mathrm{C}}^2 / \sum_i w_{i\mathrm{C}}\right]}\right\}, \tag{11}$$

$w_{i\mathrm{C}} = 1/(t^2(\mathrm{CA}) + s_i^2)$ and m_{w}(CA) is the Cochran ANOVA estimate for μ. By substituting t^2(CA2) for t^2 in Eq. (2) we get the corresponding two-step estimate, m_{w}(CA2), for μ and its approximate standard error, s.e.(m_{w}(CA2)), as defined by Eq. (3).

2.2.6. Two-step estimate of μ starting with the DerSimonian and Laird estimate for τ^2

If we substitute $a_i = 1/(t^2(\mathrm{DL}) + s_i^2)$ for $i = 1,\ldots, k$ in Eq. (6), we get the two-step estimate, t^2(DL2), where

$$t^2(\mathrm{DL2}) = \max\left\{0, \frac{\left[\sum_i w_{i\mathrm{D}}(y_i - m_{\mathrm{w}}(\mathrm{DL}))^2\right] - \left[\sum_i w_{i\mathrm{D}} s_i^2 - \sum_i w_{i\mathrm{D}}^2 s_i^2 / \sum_i w_{i\mathrm{D}}\right]}{\left[\sum_i w_{i\mathrm{D}} - \sum_i w_{i\mathrm{D}}^2 / \sum_i w_{i\mathrm{D}}\right]}\right\}, \tag{12}$$

$w_{i\mathrm{D}} = 1/(t^2(\mathrm{DL}) + s_i^2)$ and m_{w}(DL) is the DerSimonian and Laird estimate for μ. By substituting t^2(DL2) for t^2 in Eq. (2) we get the corresponding two-step estimate, m_{w}(DL2), for μ and its approximate standard error, s.e.(m_{w}(DL2)), as defined by Eq. (3).

2.2.7. Special case when sampling variances are all equal

When the intra-study sampling variance estimates $s_1^2,\ldots, s_k^2$ are approximately equal and each is replaced with their average $s^2 = (1/k)\sum_i s_i^2$, then all five estimates t^2(CA), t^2(DL), t^2(CA2), t^2(DL2), and t^2(PM) for τ^2 reduce to

$$t^2 = \max\left\{0, \frac{1}{k-1}\sum_i (y_i - y_{\mathrm{A}})^2 - s^2\right\}. \tag{13}$$

In that case the estimate for the overall treatment effect μ is the arithmetic mean $y_{\mathrm{A}} = (1/k)\sum_i y_i$ and its approximate standard error as defined by Eq. (3) is s.e.(m_{w}) $= [(1/k)(t^2 + s^2)]^{1/2}$ where $s^2 = (1/k)\sum_i s_i^2$. This expression emphasizes that s.e.(m_{w}) consists of two components of variance: an inter-study variance t^2 and an intra-study variance s^2 which reflects the individual study sampling variances s_i^2.

2.2.8. Statistical optimality of the Paule and Mandel estimate for τ^2

The Paule and Mandel estimate of the inter-study variance, t^2(PM), does not require a normality assumption. Although Paul and Mandel [3] did not evaluate the statistical properties of t^2(PM) in the original article, Rukhin et al. [4] more recently investigated its properties under normality. In particular, Rukhin et al. [4] show that when the normally assumption holds and a weighted mean of the form (1) is used as an estimate for the parameter μ, then the Paule and Mandel estimate t^2(PM) is the conditionally restricted maximum likelihood (REML) estimate of τ^2; the condition being that the estimates $s_1^2,\ldots, s_k^2$ be regarded as the true intra-study variances $\sigma_1^2,\ldots, \sigma_k^2$, respectively. A REML estimate of a variance component is an improvement over the maximum likelihood (ML) estimate because it accounts for the loss in degrees of freedom resulting from the estimation of μ [9]. Rukhin et al. [4] also show that the estimate of μ determined by using t^2(PM) for τ^2 is an approximate generalized Bayes estimate based on non-informative prior distributions for the statistical parameters μ, τ^2, and $\sigma_1^2,\ldots, \sigma_k^2$. Thus under normality, t^2(PM) and m_{w}(PM) are statistically optimal estimates for τ^2 and μ, respectively.

Under the normality assumption, the Paule and Mandel's approach is statistically optimal, but the method itself does not require a normality assumption; hence, when normality assumptions do not hold, the Paule and Mandel method is more robust for estimating τ^2 than the methods of Cochran or DerSimonian and Laird which are based on large sample assumptions.

3. Database

We compare the methods discussed in the previous section using data from six reviews published from 1981 to 2003. These reviews include articles from the New England Journal of Medicine [10], the British Medical Journal [11], Lancet [12], the Journal of the American Medical Association [13,14] and Hepatology [15]. We briefly describe the six reviews identifying each by its first author or the study name:

Baum [10]: This is a survey of a number of studies that evaluate the efficacy of antibiotics in the prevention of wound infection following colon surgery. This review includes twelve studies published before 1976 and fourteen studies published after 1976. We consider the combined data from all twenty-six studies to evaluate antibiotic prophylaxis in colon surgery.

Anand [13]: This is a meta-analysis of studies that evaluate the efficacy of long-term oral anticoagulant (OA) therapy in patients with coronary artery disease. The review includes several comparisons where the studies are stratified according to the intensity of OA therapy. We consider the nine studies that compare high-intensity OA therapy with a control group.

CLASP [12]: This is a large collaborative multi-center study of a low-dose aspirin for the prevention and treatment of pre-eclampsia in pregnant women. The review includes several meta-analyses of small and large studies to assess if aspirin is effective in preventing pre-eclampsia or perinatal deaths. We consider the six larger studies (each having more than 200 subjects) evaluating the efficacy of aspirin in reducing the incidence of pre-eclampsia.

Bradley [14]: This is a meta-analysis of four studies that evaluate the efficacy of cardiac resynchronization for the treatment of patients with advanced heart failure. The review considers several outcome measures including heart failure mortality, heart failure hospitalization, and all-cause mortality. We consider the three studies with data to assess the efficacy of cardiac resynchronization in reducing heart failure hospitalizations.

Llovet [15]: This is a systematic review of fourteen studies that evaluate the efficacy of tamoxifen (seven studies) or arterial embolization/chemoembolization (seven studies) for improving survival in patients with unresectable hepatocellular carcinoma. The review includes several comparisons evaluating 1 or 2 years survival as well as several sensitivity analyses. We consider the seven studies with data to assess if arterial embolization improves 1-year survival rates.

Teo [11]: This is an overview of seven studies that evaluate the efficacy of intravenous magnesium in suspected acute myocardial infarction. We consider all seven studies with data to assess if intravenous magnesium reduces mortality in patients with acute myocardial infarction.

Table 2 summarizes the methods and reported results from the original reviews. Except for one review [10] that uses the difference in proportions as a measure of the treatment effect, the other five reviews use the odds ratio. For estimating the overall population treatment effect μ, two reviews pool the raw data; two reviews use the modified Mantel–Haenszel (M–M-H) method [16]; and two reviews use the DerSimonian and Laird (D and L) method [1]. Columns 5 and 6 of Table 2 display the estimates of overall population treatment effect and the associated 95% confidence intervals (95% CIs) as reported in the original reviews. For each review, the estimates imply a beneficial effect of the treatment relative to control. For instance, Baum et al. [10] reported antibiotics to be highly effective in preventing wound infection following colon surgery. Their estimate for the overall difference in proportions (95% CI) of subjects who did not

Table 2
Methods and reported results from the original reviews

Study	Sample size	Outcome measure	Estimation method	Overall estimate	95% CI	Test of homogeneity (*df*)*
Baum		Difference	Pooled			
After 1976	1033	in Proportions		0.26	(0.20–0.31)	**
Before 1976	1019			0.14	(0.08–0.20)	
Anand	8749	Odds ratio	M–M-H[a]	0.52	(0.40–0.68)	**
CLASP	14530	Odds ratio	Pooled	0.81	(0.71–0.93)	10.5 (5)***
Bradley	1496	Odds ratio	D&L[b]	0.71	(0.53–0.96)	0.4 (2)
Llovet	545	Odds ratio	D&L[b]	0.64	(0.41–1.01)	7.7 (6)
Teo	1301	Odds ratio	M–M-H[a]	0.45	(0.28–0.71)	7.6 (6)

*Degrees of freedom.
**Test statistic for assessing homogeneity not reported.
***p-value < 0.10.
[a] Modified Mantel–Haenszel method [16].
[b] DerSimonian and Laird method [1].

develop infection was 0.26 (0.20–0.31) (i.e., less infections in the antibiotics group) in the studies published after 1976 and 0.14 (0.08–0.20) in the studies published before 1976. Similarly, each of the other reviews report a reduction in the odds and a benefit due to treatment with odds ratio estimates ranging from 0.45 to 0.81. Four of the reviews provide a quantitative assessment of the homogeneity of treatment effects on the odds ratio scale. Column 7 in Table 2 displays the reported test statistic for homogeneity of treatment effects. Only one study (CLASP) indicates significant heterogeneity at the 0.10 level of significance. Two reviews, Baum [10] and Anand [13], refer to effect homogeneity but do not provide a quantitative assessment. Anand [13] simply mentions lack of heterogeneity. Baum et al. [10] estimate the variability in treatment differences for each publication period (before and after 1976) using the method of Gilbert et al. [17] and conclude that the between study variation is negligible relative to the within study variation which is assumed to be equal for all studies. The combined set of all 26 studies is heterogeneous (p-value < 0.05) because the estimates of overall treatment effect for pre and post 1976 publication years are quite different.

4. Comparison

To compare the five statistical estimates for τ^2 and μ discussed in the Methods section, we use the reported data from the six reviews discussed in the previous section. We recognize that this set of reviews may not be representative of all meta-analyses published in the medical literature. In each review, the data from the i-th study consist of the total number of subjects in the treatment and control groups (n_{Ti} and n_{Ci}) and the proportion of subjects with some event of interest in each of the two groups (r_{Ti} and r_{Ci}). In this setting, several measures of treatment effect can be considered, including the difference in proportions, the odds ratio, and the relative risk [18]. To compare various estimates for τ^2 and μ, we concentrate throughout on log odds ratio as the outcome measure of interest due to its prominence in clinical trials and since it is the measure used in five of the six reviews. When the observed effect from the i-th study, y_i, is the log odds ratio, $\ln[r_{Ti}(1-r_{Ci})/r_{Ci}(1-r_{Ti})]$, we approximate the sampling variance in that study by $s_i^2=[n_{Ti}r_{Ti}(1-r_{Ti})]^{-1}+[n_{Ci}r_{Ci}(1-r_{Ci})]^{-1}$ [1].

Table 3 presents the estimates t for the inter-study standard deviation τ and Table 4 presents the estimates m_w for the overall population treatment effect μ on the log odds ratio scale determined from the five statistical methods. Table 4 also includes approximate estimates s.e.(m_w) for the standard errors of the estimates for μ. As noted earlier, these approximate estimates of standard errors are underestimates. The five statistical methods of estimation yield slightly different estimates for τ and μ in all reviews except one. For the review by Anand, the estimates for τ from all five methods are zero; therefore, the estimates for μ are the same regardless of the statistical estimation method used (Tables 3 and 4). For the other reviews, the estimates for τ from the five statistical methods are somewhat different; therefore, the estimates for μ also vary from each other (Tables 3 and 4). For example, for the review by Teo, the Cochran ANOVA estimate for τ is zero while the other estimates are greater than zero. As the estimate of τ increases, the weights assigned to the studies become more uniform, i.e., the weighted mean m_w converges to an arithmetic mean. Consequently, the larger relative impact of larger studies on the estimate of the overall treatment effect μ is reduced. Also, a larger estimate for τ yields a larger approximate estimate s.e.(m_w) of the standard error of m_w. The overall results of Tables 3 and 4 indicate that for estimation of τ, μ, and s.e.(m_w), the two-step estimates approximate the iterative

Table 3
Estimates of the inter-study standard deviation τ: log odds ratio scale

	t(PM)[a]	t(CA)[b]	t(DL)[c]	t(CA2)[d]	t(DL2)[e]
Baum	0.4796	0.4497	0.4442	0.4775	0.4771
Anand	0.0000	0.0000	0.0000	0.0000	0.0000
CLASP	0.3681	0.4410	0.2323	0.3831	0.3254
Bradley	0.0685	0.1281	0.0625	0.0793	0.0676
Llovet	0.2970	0.3284	0.2884	0.2984	0.2966
Teo	0.3312	0.0000	0.4135	0.4135	0.2883

[a] Paule and Mandel estimate.
[b] Cochran ANOVA estimate.
[c] DerSimonian and Laird estimate.
[d] Two-step estimate starting with the Cochran ANOVA estimate.
[e] Two-step estimate starting with the DerSimonian and Laird estimate.

Table 4
Estimates of the overall population treatment effect μ and the corresponding approximate standard error (in parenthesis): log odds ratio scale*

	m_w(PM) [a]	m_w(CA) [b]	m_w(DL) [c]	m_w(CA2) [d]	m_w(DL2) [e]
Baum	−1.0653 (0.1568)	−1.0614 (0.1526)	−1.0606 (0.1518)	−1.0650 (0.1565)	−1.0650 (0.1564)
Anand	−0.5742 (0.1333)	−0.5742 (0.1333)	−0.5742 (0.1333)	−0.5742 (0.1333)	−0.5742 (0.1333)
CLASP	−0.3811 (0.2060)	−0.4035 (0.2327)	−0.3240 (0.1540)	−0.3861 (0.2115)	−0.3655 (0.1901)
Bradley	−0.3398 (0.1520)	−0.3464 (0.1650)	−0.3392 (0.1511)	−0.3408 (0.1539)	−0.3397 (0.1519)
Llovet	−0.4476 (0.2225)	−0.4459 (0.2292)	−0.4481 (0.2208)	−0.4476 (0.2228)	−0.4477 (0.2224)
Teo	−0.7866 (0.3124)	−0.7533 (0.2649)	−0.8032 (0.3336)	−0.8032 (0.3336)	−0.7788 (0.3023)

*The negative estimates m_w imply that the estimate of odds ratio is less than one and the effect of treatment is beneficial.
[a] Based on t(PM), the Paule and Mandel estimate of τ.
[b] Based on t(CA), the Cochran ANOVA estimate of τ.
[c] Based on t(DL), the DerSimonian and Laird estimate of τ.
[d] Based on t(CA2), the two-step method starting with the Cochran ANOVA estimate of τ.
[e] Based on t(DL2), the two-step method starting with the DerSimonian and Laird estimate of τ.

estimates of Paule and Mandel better than the earlier one-step methods; the Paule and Mandel estimates being statistically optimal under normality.

5. Summary

We discuss two non-iterative methods and one iterative method for estimating the inter-study component of variance in a random-effects model for meta-analysis of clinical studies. We show that all three methods are special cases of a general method-of-moments estimate for the inter-study variance with slightly different weights assigned to the studies. In Cochran's ANOVA method, each study is assigned an equal weight while the weights in the DerSimonian and Laird method are inversely proportional to the within-study sampling variances. These two methods are non-iterative. The weights in the iterative Paule and Mandel method are inversely proportional to the total variances and the method requires a simple iteration that can be easily done on a spreadsheet program like Microsoft Excel; such was not the case in 1982 when the method was originally proposed and may have been a deterrent to its widespread use. We identify this approach as an extension of the same principle used with the non-iterative methods and provide the iterative algorithm to make its use more accessible in the context of clinical trials. Under normality assumptions, the Paule and Mandel estimate for the inter-study variance component is statistically optimal in the sense of being a conditionally restricted maximum likelihood estimate and the corresponding estimate for the overall treatment effect is statistically optimal in the sense of being approximately Bayes estimate. Based on the general method-of-moments estimate, we suggest two-step alternatives to the one-step non-iterative procedures based on Cochran's ANOVA and the DerSimonian and Laird methods. We compare the estimates based on these two-step procedures with those from the optimal method of Paule and Mandel. Our results indicate that the two-step methods approximate the method of Paule and Mandel better than the earlier one-step methods. The two-step methods may be considered when a non-iterative method is desired.

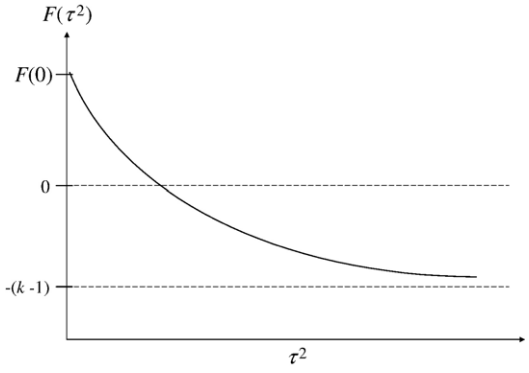

Fig. A.1. Sketch of Paule and Mandel's estimating equation when $F(0)$ is positive.

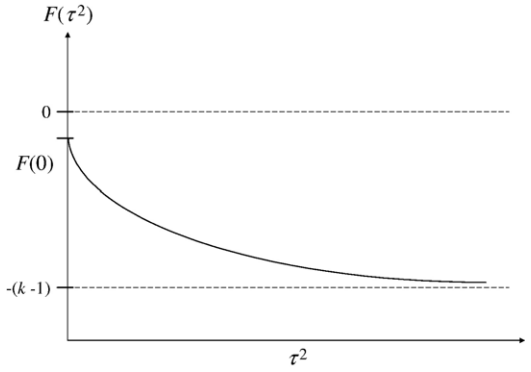

Fig. A.2. Sketch of Paule and Mandel's estimating equation when $F(0)$ is negative.

Acknowledgments

We are grateful to Katherine Halvorsen and Fern Hunt for their critical readings of the various drafts of this article and for many insightful comments.

Certain software is identified in this paper for illustration. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology or the National Institutes of Health, nor is it intended to imply that the software is necessarily the best available for the purpose.

Appendix A

The function $F(\tau^2)$ defined in Eq. (10) is strictly decreasing and concave up [8]. The shape of the function $F(\tau^2)$ is sketched in Figs. A.1 and A.2. The maximum of $F(\tau^2)$ occurs at $\tau^2=0$ and $F(\tau^2)\to-(k-1)$ as $\tau^2\to\infty$. When $F(0)$, i.e., the value of $F(\tau^2)$ at $\tau^2=0$ is positive as shown in Fig. A.1, then by the intermediate value theorem of calculus a value of τ^2 exists for which $F(\tau^2)=0$. Since $F(\tau^2)$ is strictly decreasing, such τ^2 is unique. When $F(0)$ is negative as shown in Fig. A.2, the Eq. (10) has no positive solution. When $F(0)$ is zero, the solution is $\tau^2=0$.

The solution τ^2(PM) of the estimating equation $F(\tau^2)=0$ can be determined through the following algorithm.

Start with τ^2(previous)=0 or with a number slightly above zero.

(i) Calculate weights $W_i=1/(\tau^2+s_i^2)$ for $i=1,\ldots, k$ and the function $F(\tau^2)$.

(ii) If $F(\tau^2)$ at $\tau^2=0$ is negative, set τ^2(PM)=0.
If $F(\tau^2(\text{previous}))=0$, set $\tau^2(\text{PM})=\tau^2(\text{previous})$.
If $F(\tau^2(\text{previous}))>0$, determine the correction

$$\Delta\tau^2=\frac{\sum_i W_i(y_i-y_W(\tau^2))^2-(k-1)}{\sum_i W_i^2(y_i-y_W(\tau^2))^2} \tag{14}$$

(iii) The next iterative value of $\tau^2$is $\tau^2(\text{next})=\tau^2(\text{previous})+\Delta\tau^2$.

(iv) Repeat (ii) and (iii) until $F(\tau^2(\text{previous}))=0$. The final value of τ^2 is τ^2(PM).

We suggest τ^2(CA) as the starting value for τ^2. This often reduces the number of iterations required unless τ^2(CA) is zero. This algorithm can be easily implemented in a spreadsheet program such as Microsoft Excel.

References

[1] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–87.
[2] DerSimonian R, Levine R. Resolving discrepancies between a meta-analysis and a subsequent large controlled trial. JAMA 1999;282:664–70.
[3] Paule RC, Mandel J. Consensus values and weighting factors. J Res Natl Bur Stand 1982;87:377–85.

[4] Rukhin AL, Biggerstaff BJ, Vangel MG. Restricted maximum likelihood estimation of common mean and the Mandel–Paule algorithm. J Stat Plan Inference 2000;83:319–30.
[5] Cochran WG. The combination of estimates from different experiments. Biometrics 1954;10:101–29.
[6] Kackar RN, Harville DA. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc 1984;79:853–62.
[7] Bohning D, Malzahn U, Dietz E, et al. Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. Biostatistics 2002;3(4):445–57.
[8] Kacker RN. Combining information from interlaboratory evaluations using a random effects model. Metrologia 2004;41:132–6.
[9] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. J Am Stat Assoc 1977;72:320–40.
[10] Baum ML, Anish DS, Chalmers TC, et al. A survey of clinical trials of antibiotics prophylaxis in colon surgery: evidence against further use of no-treatment controls. NEJM 1981;305:795–9.
[11] Teo K, Yusuf S, Collins R, Held P, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomized trials. BMJ 1991;303:1499–503.
[12] CLASP. A randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. Lancet 1994;343:619–29.
[13] Anand SS, Yusuf S. Oral anticoagulant therapy in patients with coronary artery disease: a meta-analysis. JAMA 1999;282:2058–67.
[14] Bradley DJ, Bradley EA, Baugham KL, et al. Cardiac resynchronization and death from progressive heart failure: a meta-analysis of randomized controlled trials. JAMA 2003;289:730–40.
[15] Llovet J, Bruix J for the Barcelona-Clinic Liver Cancer Group. Systematic review of randomized trials for unresectable hepatocellular carcinoma: chemoembolization improves survival. Hepatology 2003;37:429–42.
[16] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985;27:335–71.
[17] Gilbert JP, McPeek B, Mosteller F. Progress in surgery and anesthesia: benefits and risks of innovative therapy. In: Bunker JP, Barnes BA, Mosteller F, editors. Costs, risks and benefit of surgery. New York: Oxford University Press; 1977. p. 124–69.
[18] Whitehead A. Meta-analysis of controlled clinical trials. New York: Wiley; 2003.