STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:2693–2710 (DOI: 10.1002/sim.1482)

Improved tests for a random effects meta-regression with a single covariate

Guido Knapp*,†,‡ and Joachim Hartung‡

Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany

SUMMARY

The explanation of heterogeneity plays an important role in meta-analysis. The random effects meta-regression model allows the inclusion of trial-specific covariates which may explain a part of the heterogeneity. We examine the commonly used tests on the parameters in the random effects meta-regression with one covariate and propose some new test statistics based on an improved estimator of the variance of the parameter estimates. The approximation of the distribution of the newly proposed tests is based on some theoretical considerations. Moreover, the newly proposed tests can easily be extended to the case of more than one covariate. In a simulation study, we compare the tests with regard to their actual significance level and we consider the log relative risk as the parameter of interest. Our simulation study reflects the meta-analysis of the efficacy of a vaccine for the prevention of tuberculosis originally discussed in Berkey *et al*. The simulation study shows that the newly proposed tests are superior to the commonly used test in holding the nominal significance level. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; meta-regression; heterogeneity; trial-specific covariate

1. INTRODUCTION

One goal of the whole meta-analysis process is to possibly combine the estimates of effect from various related trials. The overall estimate is usually calculated in the random-effects model of meta-analysis [2]. This random-effects model includes a between-trial component of variance which is also called heterogeneity. The methods used in this model have been extensively discussed from both the classical and the Bayesian perspective [3–5].

However, besides the inclusion of a between-trial variance term in the analysis it is an important task to understand the possible causes of heterogeneity. One possibility is to incorporate trial-specific covariates in the model, which can then be called the random effect meta-regression. Recently, Thompson and Sharp [6] carried out a comparison of methods in

* Correspondence to: Guido Knapp, Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany.
† E-mail: knapp@statistik.uni-dortmund.de
‡ Also Project 'Meta-analysis in biometry and epidemiology' (SFB 475) of the German Research Community (DFG).

Received August 2001
Accepted December 2002

the latter model for both the 'classical' weighted least-squares approach and the Bayesian approach.

The aim of the present paper is to improve the tests on the regression parameters within the framework of the weighted least-squares approach. In the fixed- or random-effects model of meta-analysis the variance of the weighted least-squares estimator of the overall treatment effect is a concave function in the variance components of the model. By inserting (nearly) unbiased estimators of the variance components and applying Jensen's inequality the feasible variance estimator on average underestimates the true variance [7]. Thus, the commonly used tests on the overall treatment effect often yield too many unjustified significant results [8, 9], that is, the commonly used confidence intervals on the overall treatment effect do not attain the nominal significance level. Hartung and Knapp [10, 11] made these circumstances evident in their simulation studies. They proposed an improved test in this situation by deriving the non-negative invariant quadratic unbiased estimator of the variance of the overall treatment effect estimator and using the t-distribution with degrees of freedom equal to the number of trials minus one as the test distribution. In simulation studies they showed that the improved tests satisfactorily hold the prescribed significance level. We now try to assign the latter proposal to the analysis in the random effects meta-regression, that is, we mainly focus on the estimation of the variances of the parameter estimators and on deriving an appropriate test distribution.

The outline of the paper is as follows. In the next section we consider clinical trials which evaluated the efficacy of the Bacillus Calmette-Guérin (BCG) vaccine for the prevention of tuberculosis as a motivating example. The data set has been already discussed by Berkey *et al.* [1]. In Section 3 we present the random effects meta-regression and develop the new test statistics in a general framework. Section 4 contains the description and the results of our simulation study where we investigate the commonly used tests and the new tests with regard to holding the prescribed significance level. As outcome measure in the simulation study we consider the logarithmic relative risk in virtue of the analysis of the example given in Berkey *et al.* [1]. In Section 5 we discuss the explicit results in the example. Finally, some concluding remarks are given.

2. EXAMPLE

For illustrative purposes we consider a data set which has already been discussed in Berkey *et al.* [1]. This data set contains the results of clinical trials which evaluated the efficacy of the Bacillus Calmette-Guérin (BCG) vaccine for the prevention of tuberculosis (TB). The results of these trials are put together in Table I.

Some trial-specific covariates may influence the true efficacy of the vaccine. As a possible influential covariate we consider the distance of each trial from the equator, which may serve as a surrogate for the presence of environmental mycobacteria that provide a certain level of natural immunity against TB. In Figure 1 the possible relation between the trial results, the estimated logarithms of the relative risks and the places where the trials were carried out are shown. For further details, especially on the choice of the trials, we refer to Berkey *et al.* [1] and the references given therein.

From Table I we observe that the sample sizes vary extremely from trial to trial. The first three trials are rather small compared to the others, and trial 8 has more subjects in its non-vaccinated group than all the other trials together. In most trials the subjects are nearly

Table I. Data from clinical trials of BCG vaccine efficacy taken from Berkey *et al.* [1].

Trial	Vaccinated		Not vaccinated		Latitude
	Disease	No disease	Disease	No disease	
1	4	119	11	128	44°
2	6	300	29	274	55°
3	3	228	11	209	42°
4	62	13536	248	12619	52°
5	33	5036	47	5761	13°
6	180	1361	372	1079	44°
7	8	2537	10	619	19°
8	505	87886	499	87892	13°
9	29	7470	45	7232	−27°
10	17	1699	65	1600	42°
11	186	50448	141	27197	18°
12	5	2493	3	2338	33°
13	27	16886	29	17825	33°

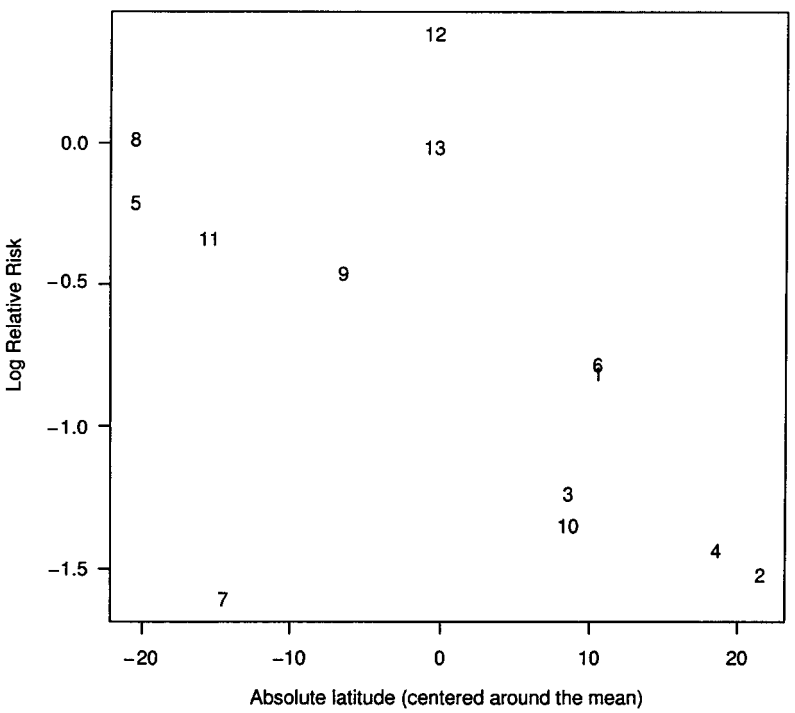

Figure 1. Log relative risk and absolute distance from the equator (centred around the mean) from 13 RCTs. The number of each trial from Table I is shown.

equally assigned to both groups. However, in trial 7 we have approximately four times more subjects in the vaccinated group than in the non-vaccinated group, and in trial 11 we have nearly twice as many subjects in the vaccinated group.

Besides the variation of the sample sizes in the trials we observe that the proportions of the diseased subjects in the non-vaccinated group differ considerably. In six trials the proportion of diseased subjects is less than 1 per cent, in trial 1 and trial 2 the proportions are 8 per cent and 9.6 per cent, respectively, and in trial 6 this proportion comes to nearly 26 per cent.

Berkey *et al.* [1] used the logarithm of the relative risk as outcome measure in their analysis. We also consider this outcome measure in our simulation study in Section 4 as well as in our analysis of the real data set in Section 5.

3. MODEL AND METHODS

The random effects meta-regression with one covariate and an additive between-trial variance [6] is given by

$$y_i \sim \mathrm{N}(\alpha + \beta x_i, \tau^2 + \xi_i), \quad i = 1, \ldots, k \tag{1}$$

that is, the observed parameter of interest y_i in trial i, $i = 1, \ldots, k$, is, at least approximately, normally distributed with expected value $\alpha + \beta x_i$ and variance $\tau^2 + \xi_i$. The value of the trial-specific covariate x_i is known, the parameter β represents the change in the parameter of interest per unit change in the covariate x_i, and α is the parameter of interest if the covariate is equal to zero. The variance of y_i consists of two components. The parameter ξ_i stands for the variance of the observed parameter of interest in trial i, also known as the within-trial variance, and τ^2 is the between-trial variance.

Let $w_i = (\tau^2 + \xi_i)^{-1}$, $\lambda_i = w_i / w$, $i = 1, \ldots, k$, with $w = \sum_{i=1}^{k} w_i$, then the weighted least-squares estimators of α and β are given by

$$\tilde{\beta} = \frac{\sum_{i=1}^{k} \lambda_i x_i y_i - \sum_{i=1}^{k} \lambda_i x_i \sum_{i=1}^{k} \lambda_i y_i}{\sum_{i=1}^{k} \lambda_i x_i^2 - \left(\sum_{i=1}^{k} \lambda_i x_i \right)^2} \tag{2}$$

and

$$\tilde{\alpha} = \sum_{i=1}^{k} \lambda_i \, y_i - \tilde{\beta} \sum_{i=1}^{k} \lambda_i x_i \tag{3}$$

The variances and covariance of the estimators (2) and (3) are

$$\mathrm{var}(\tilde{\alpha}) = \left(\sum_{i=1}^{k} w_i - \left(\sum_{i=1}^{k} w_i x_i \right)^2 \Big/ \sum_{i=1}^{k} w_i x_i^2 \right)^{-1} \tag{4}$$

$$\mathrm{var}(\tilde{\beta}) = \left(\sum_{i=1}^{k} w_i x_i^2 - \left(\sum_{i=1}^{k} w_i x_i \right)^2 \Big/ \sum_{i=1}^{k} w_i \right)^{-1} \tag{5}$$

and

$$\mathrm{cov}(\tilde{\alpha}, \tilde{\beta}) = \frac{-\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i \sum_{i=1}^{k} w_i x_i^2 - \left(\sum_{i=1}^{k} w_i x_i \right)^2} \tag{6}$$

To apply the above formulae estimates of the variance components τ^2 and ξ_i, $i=1,\ldots,k$, have to be put in. Usually, every trial i produces an estimate of ξ_i. The between-trial variance can be estimated using either a method of moments estimator, a restricted maximum likelihood estimator, or an empirical Bayes estimator [6] which are briefly described below. In the following, with $\hat{w}_i = 1/(\hat{\tau}^2 + \hat{\xi}_i)$, $i=1,\ldots,k$, we denote the estimate of w_i, and $\hat{\alpha}$ and $\hat{\beta}$ are the feasible weighted least-squares estimators of α and β with weights $\hat{w}_i$.

The method of moments (MM) estimator of the between-trial variance τ^2 can be derived from the heterogeneity statistic $Q_1 = \sum_{i=1}^{k} w_i^*(y_i - \hat{\alpha}^* - \hat{\beta}^* x_i)^2$, where $\hat{\alpha}^*$ and $\hat{\beta}^*$ are the feasible weighted least-squares estimator with weights $w_i^* = \hat{\xi}_i^{-1}$, $i=1,\ldots,k$ [12]. In model (1) the method of moments estimator is given in its truncated form as

$$\hat{\tau}^2_{\mathrm{MM}} = \max\left\{0, \frac{Q_1 - (k-2)}{F(w^*, x)}\right\} \tag{7}$$

with

$$F(w^*, x) = \sum w_i^* - \frac{\sum w_i^{*2} \sum w_i^* x_i^2 - 2\sum w_i^{*2} x_i \sum w_i^* x_i + \sum w_i^* \sum w_i^{*2} x_i^2}{\sum w_i^* \sum w_i^* x_i^2 - (\sum w_i^* x_i)^2}$$

In the case of more than one covariate the method of moments estimator can easily be extended. The resulting formula is given in Appendix A1. When there is no covariate the method of moment estimator reduces to the DerSimonian–Laird estimator [2].

The (approximate) restricted maximum likelihood (REML) estimator [13] for the between-trial variance in model (1) with one covariate is the solution of the equation

$$\tau^2 = \frac{\sum_{i=1}^{k} \hat{w}_i^2((k/(k-2))(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 - \hat{\xi}_i)}{\sum_{i=1}^{k} \hat{w}_i^2} \tag{8}$$

This equation is iteratively solved using a starting value of τ^2, say $\tau^2 = \tau_0^2$. With the weights $\hat{w}_i = 1/(\tau_0^2 + \hat{\xi}_i)$ the initial values of $\hat{\alpha}$ and $\hat{\beta}$ are given. Then, the right hand side of (8) can be evaluated to yield a new value of τ^2. This provides new weights $\hat{w}_i$, and leads to new estimates of α and β and finally to a new value of τ^2. The procedure continues until convergence under the restriction that τ^2 is non-negative.

Berkey *et al.* [1] uses the empirical Bayes (EB) estimator [13] of the between-trial variance which is also iteratively given as the solution of the equation

$$\tau^2 = \frac{\sum_{i=1}^{k} \hat{w}_i((k/(k-2))(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 - \hat{\xi}_i)}{\sum_{i=1}^{k} \hat{w}_i} \tag{9}$$

Note that the difference between equations (8) and (9) is that $\hat{w}_i^2$ in (8) is replaced by $\hat{w}_i$ in (9). Equation (9) can be solved in the same manner as described for the solution of equation (8).

The commonly used $(1-\kappa)$-confidence intervals on the parameters α and β are then given by

$$\hat{\alpha} \pm \sqrt{\{\widehat{\mathrm{var}}(\hat{\alpha})\}} c_{1-\kappa/2} \tag{10}$$

and

$$\hat{\beta} \pm \sqrt{\{\widehat{\mathrm{var}}(\hat{\beta})\}} c_{1-\kappa/2} \tag{11}$$

where $\widehat{\text{var}}(\hat{\alpha})$ and $\widehat{\text{var}}(\hat{\beta})$ are given by putting $\hat{w}_i,\ i=1,\ldots,k$, in (4) and (5), respectively, and $c_{1-\kappa/2}$ denotes the $(1-\kappa/2)$-quantile of an appropriate statistical distribution like the standard normal distribution or the t-distribution with v degrees of freedom. Berkey *et al.* [1] proposed using the t-distribution with $(k-5)$ degrees of freedom in (10) and (11) based on an empirical result of their simulation study. Note that the use of the t-distribution with a decreasing number of degrees of freedom always increases the actual confidence coefficient of the intervals but one has no theoretical foundation of the appropriate choice of the number of degrees of freedom. In the simulation study, presented in Section 4, we will use the t-distribution with $(k-2)$ degrees of freedom in regard to the derivation of our newly proposed confidence interval described below.

Moreover, the confidence intervals (10) and (11) can be used to test the two-sided hypotheses $H_0{:}\alpha=\alpha_0$ versus $H_1{:}\alpha\neq\alpha_0,\ \alpha_0$ given, and $H_0{:}\beta=\beta_0$ versus $H_1{:}\beta\neq\beta_0,\ \beta_0$ given, respectively. The null hypotheses is rejected if the corresponding confidence interval does not contain the parameter value under the null hypothesis.

In the case of no covariate, the above confidence interval on α does not yield satisfactory results because it is often too narrow when the standard normal distribution or the t-distribution with $(k-1)$ degrees of freedom is used as a test distribution, except for relatively small values of τ^2 where the confidence intervals can become conservative by using the t-distribution. Hartung and Knapp [10, 11] proposed an improved confidence interval in this situation by considering a refined variance estimator. Their interval is an exact $(1-\kappa)$-confidence interval for known variance components and yields good results concerning the actual confidence coefficient by replacing the usually unknown variance components through appropriate estimates. Now, we carry over their approach to the analysis in the meta-regression (1).

Let us consider the quadratic form

$$q=\frac{1}{k-2}\sum_{i=1}^{k}w_i(y_i-\tilde{\alpha}-\tilde{\beta}x_i)^2 \tag{12}$$

The quadratic form q can be seen as a mean sum of the weighted least-squares residuals. In Appendix A2 it is shown that, under normality of y, the quadratic form q from (12) is stochastically independent of the weighted least-squares estimators $\tilde{\alpha}$ and $\tilde{\beta}$, and that $(k-2)q$ is χ^2-distributed with $k-2$ degrees of freedom. Consequently, the expected value of q is equal to one for known variance components and under normality assumption.

Hence, unbiased and non-negative estimators of the variances of $\tilde{\alpha}$ and $\tilde{\beta}$ are given by

$$q(\tilde{\alpha})=\frac{1}{k-2}\sum_{i=1}^{k}g_i(y_i-\tilde{\alpha}-\tilde{\beta}x_i)^2 \tag{13}$$

with $g_i=w_i/[\sum w_i-(\sum w_ix_i)^2/\sum w_ix_i]^2,\ i=1,\ldots,k$, and

$$q(\tilde{\beta})=\frac{1}{k-2}\sum_{i=1}^{k}h_i(y_i-\tilde{\alpha}-\tilde{\beta}x_i)^2 \tag{14}$$

with $h_i=w_i/[\sum w_ix_i^2-(\sum w_ix_i)^2/\sum w_i],\ i=1,\ldots,k$.

Replacing the unknown variance components in (13) and (14) by appropriate estimates, we propose the following confidence intervals on α and β

$$\hat{\alpha}\pm\sqrt{\{\hat{q}(\hat{\alpha})\}}t_{k-2,1-\kappa/2} \tag{15}$$

and

$$\hat{\beta} \pm \sqrt{\{\hat{q}(\hat{\beta})\}} t_{k-2,1-\kappa/2} \tag{16}$$

where $t_{v;\kappa}$ denotes the κ-quantile of the t-distribution with v degrees of freedom.

Note that, with the use of the EB estimator of τ^2, we obtain that the mean sum of the weighted least-squares residuals, that is, q from (12), is equal to one when the solution of the estimation equation (9) is positive. This can be seen by rearranging equation (9). Hence, the confidence intervals (10) and (15) on α as well as the confidence intervals (11) and (16) on β are identical given a positive EB between-trial variance estimate and equal test distribution. Furthermore, with the use of the MM estimator or the REML estimator of the between-trial variance, the confidence intervals (15) and (16) are smaller than the corresponding intervals (10) and (11) when the realized value of the quadratic form q from (12) is less than one given equal test distribution in both cases. Therefore, we may consider an *ad hoc* modification of the feasible variance estimates $\hat{q}(\hat{\alpha})$ and $\hat{q}(\hat{\beta})$ in the limits of the confidence intervals (15) and (16) to the effect that we force the realized value of q to be greater or equal to one. That is, we proposed the modified confidence intervals

$$\hat{\alpha} \pm \sqrt{\{\hat{q}^*(\hat{\alpha})\}} t_{k-2,1-\kappa/2} \tag{17}$$

with

$$\hat{q}^*(\hat{\alpha}) = \frac{1}{\sum \hat{w}_i - (\sum \hat{w}_i x_i)^2 / \sum \hat{w}_i x_i^2} \max\left\{1; \frac{1}{k-2} \sum \hat{w}_i (y_i - \hat{\alpha} - \hat{\beta} x_i)^2\right\}$$

and

$$\hat{\beta} \pm \sqrt{\{\hat{q}^*(\hat{\beta})\}} t_{k-2,1-\kappa/2} \tag{18}$$

with

$$\hat{q}^*(\hat{\beta}) = \frac{1}{\sum \hat{w}_i x_i^2 - (\sum \hat{w}_i x_i)^2 / \sum \hat{w}_i} \max\left\{1; \frac{1}{k-2} \sum \hat{w}_i (y_i - \hat{\alpha} - \hat{\beta} x_i)^2\right\}$$

4. SIMULATION STUDY

In a simulation study we compare the actual levels of the confidence intervals, or equivalently of the corresponding hypothesis tests, discussed in the last section. We choose the logarithmic relative risk as the parameter of interest. The random numbers in the simulation study are generated using the SAS functions RANNOR (normal distribution) and RANBIN (binomial distribution). The 2×2 tables are generated according to the principle discussed in Berkey *et al.* [1]. Let p_c denote the probability of disease among the control subjects, let α and β be given, let $\delta_i \sim N(0, \tau^2)$ be a normal variate with mean zero and variance τ^2, τ^2 also given, then the probability of disease in the vaccinated group of trial i is calculated as $p_{vi} = p_c \exp(\alpha + \beta(x_i - \bar{x}) + \delta_i)$. Finally, the number of diseased subjects a_i in the vaccinated group in trial i is randomly selected from a binomial variate with parameters n_{vi} and p_{vi}, and the number of diseased subjects c_i in the control group in trial i is randomly generated from a binomial variate with parameters n_{ci} and p_c.

The nominal significance level is chosen as $\kappa = 0.05$. Each estimated significance level is based on 10 000 replications of the respective model, that is, for estimates between 0.0458 and 0.0542 the corresponding 95 per cent confidence intervals contain the value of the nominal level $\kappa = 0.05$.

The choice of the parameters in the simulation study ensues from the results in the example presented in Section 2. As values of the probability of disease in the control group we consider $p_c = 0.05$ and $p_c = 0.1$. The other parameters are set to $\alpha = -0.5$ and $\beta = -0.02$, and as the values of the between-trial variance we take $\tau^2 = 0, 0.05, 0.1, 0.2$ and 0.3.

In the first part of the simulation study we use the same arrangement as in the example from Section 2, that is, we consider $k = 13$ trials with the given sample sizes n_{vi} (vaccinated group) and n_{ci} (control group), $i = 1, \ldots, k$, and as the covariate we take the values of the absolute latitude centred about the mean latitude of all trials as described in Berkey *et al.*

In the second part, we extend the simulation study to check the generalization of our results from the first part. We consider balanced trials, that is, the sample sizes in the treatment group and in the control group are identical. As possible values of the sample sizes we choose the sample sizes from the vaccinated group in the example discussed throughout this paper. The number of trials are $k = 5, 7, 10$ and 15. In each run of the simulation a sample of k values from the considered pool of sample sizes is drawn with replications and each drawn value is the sample size of the treatment group as well as of the control group for a trial. The vector of the covariate is randomly assigned with random numbers from a standard normal distribution in each run.

The estimated log relative risk [14] is calculated in every trial as

$$y_i = \log(\mathrm{RR}_i) = \log\left(\frac{a_i + 0.5}{n_{vi} + 0.5}\right) - \log\left(\frac{c_i + 0.5}{n_{ci} + 0.5}\right), \quad i = 1, \ldots, k \tag{19}$$

and the estimated variance of y_i in every trial has the form

$$\hat{\xi}_i = \widehat{\mathrm{var}}(\log(\mathrm{RR}_i)) = \frac{1}{a_i + 0.5} - \frac{1}{n_{vi} + 0.5} + \frac{1}{c_i + 0.5} - \frac{1}{n_{ci} + 0.5} \tag{20}$$

In the following, we call the variance estimator from (20) the 'usual' within-trial variance estimator.

The estimator of the log relative risk y_i and its variance estimator $\hat{\xi}_i$ are correlated, and Berkey *et al.* [1] reports that the estimates of the parameters α, β, and τ^2 are slightly biased towards zero in their simulation study. Thus, they proposed a smoothed estimator of the within-trial variance which reduces the correlation between $\log(\mathrm{RR}_i)$ and $\widehat{\mathrm{var}}(\log(\mathrm{RR}_i))$. We adopt this estimator in a slightly modified manner as

$$\hat{\xi}_i^* = \frac{1}{n_{vi}} \frac{1}{k} \sum_{i=1}^{k} \frac{n_{vi} - a_i + 0.5}{a_i + 0.5} + \frac{1}{n_{ci}} \frac{1}{k} \sum_{i=1}^{k} \frac{n_{vi} - c_i + 0.5}{c_i + 0.5} \tag{21}$$

In the following, the variance estimator (21) will be denoted as the 'smoothed' within-trial variance estimator.

4.1. Results in the model with the setting of the example

First, we report the results on the goodness of the between-trial estimators discussed in Section 3. For the solution of the REML equation (8) and the EB equation (9), respectively, we

Table II. Means and standard deviations (in parentheses) of the three various between-trial variance estimators in dependence of the choice of the within-trials variance estimates given a probability of disease in the control group of $p_c = 0.05$.

τ^2	Usual within-trial variance estimators			Smoothed within-trial variance estimators		
	MM	REML	EB	MM	REML	EB
0	0.00079 (0.00140)	0.00026 (0.00114)	0.00268* (0.00764)	0.00141 (0.00266)	0.00054 (0.00208)	0.00889* (0.03206)
0.05	0.04901 (0.04345)	0.04586 (0.02961)	0.04952† (0.03400)	0.05014† (0.04273)	0.04919 (0.03552)	0.06368 (0.06187)
0.1	0.09681 (0.08216)	0.09301 (0.05311)	0.09709 (0.05715)	0.09902† (0.08125)	0.09977† (0.06130)	0.11354 (0.08148)
0.2	0.19505 (0.16492)	0.18905 (0.09948)	0.19281 (0.10286)	0.20185† (0.16741)	0.20246 (0.11153)	0.21340 (0.12500)
0.3	0.28845 (0.24357)	0.28178 (0.14030)	0.28620 (0.14411)	0.30076† (0.25129)	0.30023† (0.15555)	0.30989 (0.16511)

*Estimates influenced by some large values.
†95 per cent confidence interval for τ^2 contains the true value.

chose as a starting value always $\tau_0^2 = 0$. The convergence was reached when the difference of the estimated between-trial variance of two successive iterations was less than 1×10^{-8} or the number of iterations was equal to 1000. In our simulation study we did not observe significant differences in the attitude of the between-trial variance estimators in dependence of the value of the probability of disease in the control group. Thus, in Table II, we put together the means and the standard deviations of the empirical distribution of the three estimators given $p_c = 0.05$ and given the type of the within-trial variance estimates.

With the 'usual' within-trial variance estimators, we observe that, for positive values of τ^2, all means indicate an underestimation of the true parameter value in average. This can be supported by calculating a 95 per cent confidence interval on τ^2 based on the central limit theorem (mean $\pm$ 1.96 $\times$ standard deviation/$\sqrt{10\,000}$). Only for $\tau^2 = 0.05$ the interval with the EB estimators contains the true between-trial variance, in all the other cases the intervals do not contain the true value. Moreover, the REML estimator has a larger downward bias than the other two estimators, while the MM estimator has a larger variation. For $\tau^2 = 0$, the MM estimator and the REML estimator overestimate the true value which is expected because of the construction. For the EB estimator the estimation equation tends to produce some larger positive values in this situation which substantially affect the mean and the standard deviation.

In comparison to the use of the 'usual' within-trial variance estimators, the between-trial variance estimators on average yield larger values when the 'smoothed' within-trial variance estimates are put in the corresponding formulae. For positive values of τ^2, the MM estimator and the REML estimator are nearly unbiased. This can be seen by calculating the confidence intervals on τ^2 in the same manner as described above. The 95 per cent confidence intervals with the MM estimator always overlap the true between-trial variance. For the REML

estimator, the intervals with true values of $\tau^2 = 0.1$ and $\tau^2 = 0.3$ contain the parameter value, while in the other two cases the true value is barely outside the intervals. However, the variation of the REML estimator is less than the variation of the MM estimator. The EB estimator, however, on average overestimates the between-trial variance and even the 99 per cent confidence intervals on τ^2 do not contain the true parameter value. Based on the just described results one may prefer to use the MM estimator or the REML estimator of the between-trial variance with the 'smoothed' within-trial variance estimates to judge the amount of unexplained heterogeneity.

Let us now consider the results of the estimated levels of the two-sided tests on $H_0: \alpha = -0.5$ and on $H_0: \beta = -0.02$ based on the different confidence intervals presented in Section 3. In the simulation study we found out that the use of the 'smoothed' within-trial variance estimators (21) improves the levels of the tests towards the nominal level of $\kappa = 0.05$ compared to the 'usual' within-trial variance estimators (20) for positive values of τ^2. Thus, we only present the results by using the 'smoothed' within-trial variance estimators. The estimated levels of the tests in dependence of the choice of the between-trial variance estimators are put together in Table III. In all tests, we use the t-distribution with $(k - 2)$ degrees of freedom as test distribution.

The results in Table III show that, for positive values of τ^2, the tests with the commonly used test statistics (10) and (11) are rather anticonservative when the MM estimator of the between-trial variance is used irrespective of the choice of τ^2 and of p_c. Using the REML estimator of τ^2 we observe that the estimated levels of the tests are always smaller than the estimated levels in the latter case and, for growing values of τ^2, the estimated levels tend to the prescribed level. For $p_c = 0.1$, the estimated sizes are in general smaller compared to $p_c = 0.05$. With the newly proposed tests (15) and (16) and their *ad hoc* modifications (17) and (18) we observe an improvement concerning the actual significance level of the tests given the nominal level of $\kappa = 0.05$ by using the MM estimator of τ^2 as well as by using the REML estimator of τ^2. Because of their construction, the tests (17) and (18) possess smaller significance levels than the corresponding tests (15) and (16). For $p_c = 0.05$, the tests are still a bit anticonservative for $\tau^2 = 0.05$, but with increasing values of the between-trial variance, especially the tests with the *ad hoc* modification attain the nominal level. For $p_c = 0.1$, the estimated levels of the tests are in general less compared to $p_c = 0.05$ given the value of τ^2. For larger values of the between-trial variance, the *ad hoc* modified tests can even become conservative, whereas for $\tau^2 = 0.05$ these tests nearly attain the prescribed level. Note that the *ad hoc* modification of the tests (17) and (18) has much more influence on the actual test size when the MM estimator of τ^2 is used compared to the use of the REML estimator. The tests using the EB estimator of τ^2 have estimated significance levels which are slightly larger than the estimated levels of the *ad hoc* modified tests with the MM estimator or the REML estimator of τ^2.

If no between-trial variance is present, that is $\tau^2 = 0$, all the commonly used tests as well as the *ad hoc* modified tests are extremely conservative irrespective of the choice of p_c. The newly proposed tests (15) and (16), however, are a bit anticonservative in this situation.

4.2. Results with different numbers of trials and arbitrary vectors of the covariate

In the second part of the simulation study we check the generalization of our results of the first part. Because of the sampling of the sample sizes described above we allow a large

Table III. Estimated levels (in per cent) of the two-sided tests on H_0:$\alpha = -0.5$ (first row) using (10), (15), (17) and on H_0:$\beta = -0.02$ (second row) using (11), (16), (18) given a significance level of $\kappa = 0.05$ in the setting of the example between-trial variance estimators.

τ^2	(10) (11)	(15) (16) MM	(17) (18)	(10) (11)	(15) (16) REML	(17) (18)	(17) (18) EB
$p_c = 0.05$							
0	2.6	5.7	2.4	3.3	5.9	2.6	2.4
	2.3	5.7	2.2	3.1	6.1	2.5	2.0
0.05	7.8	6.7	5.5	7.4	7.0	6.3	6.1
	8.1	7.2	6.1	8.0	7.5	6.7	6.6
0.1	7.8	6.2	4.9	6.4	5.8	5.4	5.4
	8.3	6.7	5.5	6.8	6.4	5.9	5.9
0.2	8.6	6.3	5.0	5.8	5.7	5.2	5.5
	8.2	6.1	5.0	5.9	5.8	5.4	5.5
0.3	8.2	5.8	4.7	5.6	5.4	5.2	5.3
	8.7	5.8	4.8	5.5	5.4	5.0	5.2
$p_c = 0.1$							
0	2.7	5.9	2.6	3.6	6.2	2.9	2.4
	2.6	5.6	2.6	3.5	6.0	2.9	2.4
0.05	8.1	6.3	5.2	6.2	5.9	5.3	5.4
	8.2	6.6	5.3	6.6	6.3	5.7	5.9
0.1	7.9	5.6	4.6	5.3	5.2	4.8	5.2
	8.4	5.9	4.8	5.6	5.5	5.0	5.3
0.2	8.3	5.6	4.5	5.3	5.2	5.0	5.2
	8.9	6.0	4.9	5.8	5.6	5.3	5.5
0.3	8.1	5.1	4.1	5.0	4.9	4.7	4.8
	8.2	5.3	4.2	5.0	5.0	4.7	4.9

variation of the actual sample sizes. Consequently, the relative weights, that is, the trial-specific inverse variance estimates divided by the total sum of all inverse variance estimates, that the trial contributes to the analysis may differ extremely. The variation or the range of the relative weights is larger for a small number of trials, that is, the consideration of $k = 5$ trials reflects the most unfavourable situation in the simulation study with respect to the actual significance level.

In Table IV the estimated actual significance levels of the hypothesis tests are put together for $p_c = 0.05$. As in the first part of the simulation study the estimated significance levels given $p_c = 0.1$ are in general closer to the pre-chosen level in the settings considered now. Moreover, we restrict the presentation on the results using the 'smoothed' within-trial variance estimators as in Table III. However, it is worthwhile noting that the results obtained by using the 'usual' or the 'smoothed' within-trial variance estimators do not differ so much in the present variable settings as in the setting of the example.

First, we observe that the newly proposed tests (15) and (16) attain the pre-chosen significance level when $\tau^2 = 0$ for all considered number of trials irrespective of the choice of the estimator of τ^2. The other tests are all rather conservative given no between-trial variance.

Table IV. Estimated levels (in per cent) of the two-sided tests on H_0:$\alpha = -0.5$ (first row) using (10), (15), (17) and on H_0:$\beta = -0.02$ (second row) using (11), (16), (18) given a significance of $\kappa = 0.05$ with $p_c = 0.05$, the vector of covariates from the standard normal distribution and different number of trials.

τ^2	(10) (11)	(15) (16) MM	(17) (18)	(10) (11)	(15) (16) REML	(17) (18)	(17) (18) EB
K = 5							
0	0.1	5.0	0.1	0.2	5.0	0.1	0.1
	0.1	5.0	0.1	0.2	5.0	0.2	0.1
0.05	7.0	10.0	6.6	10.6	12.1	9.0	6.9
	6.2	8.9	5.8	9.2	10.7	7.8	6.1
0.1	8.3	9.9	7.4	11.5	12.1	10.1	7.8
	7.1	8.7	6.5	10.2	10.5	8.7	6.8
0.2	7.8	8.6	6.7	10.5	10.4	9.1	7.1
	7.5	8.0	6.3	10.1	9.5	8.4	6.6
0.3	7.5	7.6	6.3	9.9	9.5	8.6	6.6
	6.9	7.0	5.8	9.2	8.8	7.8	5.9
K = 7							
0	0.7	5.0	0.7	0.9	5.1	0.8	0.7
	0.8	4.9	0.8	1.0	5.0	0.8	0.8
0.05	8.1	8.9	7.1	10.5	10.7	9.3	7.6
	7.6	8.4	6.7	9.7	9.9	8.4	7.0
0.1	8.0	7.7	6.4	9.3	9.0	8.1	7.0
	7.8	7.8	6.3	9.2	8.7	7.9	7.0
0.2	8.4	7.8	6.5	8.9	8.6	8.0	7.1
	8.2	7.2	6.1	8.2	7.8	7.3	6.4
0.3	7.9	7.0	5.8	7.4	7.2	6.7	6.1
	8.1	6.9	5.6	7.4	7.3	6.6	6.3
K = 10							
0	1.6	4.6	1.6	1.9	4.7	1.6	1.6
	1.8	5.0	1.8	2.2	5.2	1.9	1.8
0.05	8.6	8.1	6.8	9.1	9.0	8.0	7.2
	7.7	7.4	6.3	8.6	8.3	7.6	6.9
0.1	7.6	6.9	5.6	7.4	7.2	6.5	6.2
	8.0	7.1	6.1	7.5	7.4	6.8	6.6
0.2	7.9	6.5	5.3	6.6	6.4	6.0	5.9
	8.1	6.5	5.6	6.8	6.4	6.0	5.9
0.3	8.4	6.4	5.3	6.5	6.3	5.9	5.8
	7.8	6.3	5.1	6.2	6.0	5.6	5.6
K = 15							
0	1.7	4.2	1.7	2.3	4.4	1.9	1.7
	1.8	4.4	1.7	2.3	4.7	2.0	1.7
0.05	7.7	6.9	5.6	7.3	7.0	6.3	6.5
	7.5	6.9	5.8	7.2	7.1	6.4	6.5
0.1	7.1	6.1	5.1	6.0	5.9	5.4	5.5
	7.4	6.1	5.0	6.0	5.9	5.4	5.6
0.2	7.9	6.2	5.1	6.1	5.9	5.5	5.7
	7.2	5.6	4.6	5.4	5.4	4.9	5.1
0.3	8.1	6.1	5.0	5.9	5.7	5.5	5.6
	8.1	5.7	4.7	5.6	5.5	5.1	5.4

Table V. Influence of the choice of the within-trial variance estimate.

Trial number	'Usual' within-trial estimate $\hat{\xi}_i$	'Smoothed' within-trial estimate $\hat{\xi}_i^*$	'Weights[†] with 'usual' estimate (in per cent)	'Weights[‡] with 'smoothed' estimate (in per cent)	Weights[§] with sample size (in per cent)
1	0.2939	2.8321	0.55	0.07	0.07
2	0.1811	1.2043	0.90	0.17	0.17
3	0.3638	1.6233	0.45	0.13	0.13
4	0.0199	0.0277	8.20	7.54	7.41
5	0.0505	0.0683	3.22	3.05	3.04
6	0.0069	0.2446	23.66	0.85	0.84
7	0.2109	0.3372	0.77	0.62	0.89
8	0.0040	0.0042	41.18	50.24	49.47
9	0.0556	0.0496	2.93	4.21	4.13
10	0.0712	0.2167	2.29	0.96	0.95
11	0.0124	0.0100	13.17	20.94	21.82
12	0.4667	0.1512	0.35	1.38	1.35
13	0.0701	0.0212	2.32	9.84	9.73

[†] $(1/\hat{\xi}_i)/\sum(1/\hat{\xi}_i) \times 100$.
[‡] $(1/\hat{\xi}_i^*)/\sum(1/\hat{\xi}_i^*) \times 100$.
[§] $(n_{ci} + n_{vi})/\sum(n_{ci} + n_{vi}) \times 100$.

For a positive between-trial variance τ^2 we can conclude from the simulation study that with an increasing number of trials the tests get actual significance levels closer to the nominal significance level and the part of Table IV with $k = 15$ trials essentially reflects the result from Table III.

For $k = 5$ trials all tests exceed the pre-chosen level and especially the use of the REML estimator of τ^2 leads to rather anticonservative tests. The *ad hoc* modification of the newly proposed tests, that is, using (17) and (18), in connection with the MM estimator or the EB estimator of τ^2 yield the best results for $k = 5$. For a larger number of trials the tests with the REML estimator of τ^2 performs better but the *ad hoc* modification of the newly proposed tests with one of the other two estimators of τ^2 are always preferable. Especially for a small number of trials the *ad hoc* modification improves the original newly proposed tests using (15) and (16).

5. RESULTS IN THE EXAMPLE

In this section we produce results in the example from Section 2 additionally to the results by Berkey *et al.* [1]. Let us first discuss the use of the within-trial variance estimates. In meta-analysis, the single estimate usually reflects the precision of a trial. In Table V we put together the results of the two different within-trial variance estimators discussed in the previous section. Moreover, we calculate the normed weight of every trial, that is, the single inverse within-trial variance estimate divided by the sum of all inverse within-trial variance estimates, under the hypothesis $\tau^2 = 0$ and the proportion of the sample size of every trial to the total sample size in the meta-analysis.

We observe that the orders of the trials based on their precision do not coincide using the two different estimators. With the 'usual' within-trial variance estimates, trial 6 has the second highest precision, but by using the 'smoothed' within-trial variance estimates this precision is not remarkable compared to the other trials. Conversely, trial 13 had a larger precision when the 'smoothed' within-trial variance estimate was used. It is worthwhile noting that the order of precision based on the 'smoothed' within-trial variance estimates is exactly the same as the order of precision simply based on the sample sizes. Note that, in the random effects meta-regression (1) with $\hat{\tau}^2 > 0$, the resulting weights $\hat{\lambda}_i = \hat{w}_i/(\sum \hat{w}_i)$ have the same order but the trials are more 'equally' weighted, that is, for instance, the strong influence of trial 8 diminishes for growing $\hat{\tau}^2$.

In the numerical meta-analysis we use the same formulae as in the simulation study. For this reason, some results slightly differ from the results given in Berkey *et al.* [1].

With the 'usual' within-trial variance estimates, the estimates of the between-trial variance are nearly identical using the MM estimator or the REML estimator ($\hat{\tau}^2 = 0.0622$ (MM), $\hat{\tau}^2 = 0.0614$ (REML)). The EB estimator yields an estimate of $\hat{\tau}^2 = 0.1388$, which is more than twice as big as the other estimates. The regression equation with the EB estimator of $\hat{\tau}^2$ is then $y_i = -0.720 - 0.028(x_i - \bar{x})$ and with the MM estimator $y_i = -0.708 - 0.029(x_i - \bar{x})$, which is nearly the same equation using the REML estimator. The effect of the newly proposed confidence intervals can be seen by calculating the t-score, that is, the estimate divided by its estimated standard error. Using the MM estimator of $\hat{\tau}^2$, the t-scores are -7.08 for α and -4.30 for β with the commonly used standard error estimates and -6.00 for α and -3.64 for β with the newly proposed standard error estimates. The t-score, using the REML estimator of $\hat{\tau}^2$, are of the same magnitude in both cases; that means, the resulting newly proposed confidence intervals are substantially wider than the commonly used ones.

By using the 'smoothed' within-trial variance estimates, the estimates of $\hat{\tau}^2$ are given by $\hat{\tau}^2 = 0.1013$ (MM), $\hat{\tau}^2 = 0.1360$ (REML) and $\hat{\tau}^2 = 0.1479$ (EB). We see that the EB estimate is only a bit larger than in the first case, whereas the other two estimates are considerably larger but they are still smaller than the EB estimate in the first case. The resulting regression equations are given by $y_i = -0.618 - 0.027(x_i - \bar{x})$ (MM), $y_i = -0.624 - 0.027(x_i - \bar{x})$ (REML), and $y_i = -0.626 - 0.026(x_i - \bar{x})$ (EB). Let us again consider the t-scores to judge the influence of the newly proposed estimates of the standard error. Using the MM estimator of $\hat{\tau}^2$, the t-scores are -4.52 for α and -2.87 for β in the commonly used tests and -4.12 for α and -2.62 for β in the newly proposed tests. The differences between the t-scores are smaller when the REML estimate is used. In the commonly used tests they are -4.15 for α and -2.56 for β and, in the newly proposed test, -4.07 for α and -2.50 for β.

6. CONCLUDING REMARKS

In this paper we proposed new standard error estimates for the regression parameter estimates of the weighted least-squares approach in the random effects meta-regression. This proposal leads to new confidence intervals and hypothesis tests, respectively, on the regression parameters. Based on some theoretical considerations, the test distribution to be used is the t-distribution where the number of degrees of freedom is determined by the number of trials minus the number of regression parameters. As possible estimators of the between-trial variance we considered a method of moments estimator, a restricted maximum likelihood

estimator, and an empirical Bayes estimator. For positive estimates of the empirical Bayes estimators we showed that the commonly used test statistics and the newly proposed test statistics coincide because the estimation equation of the empirical Bayes estimator forces the mean sum of the weighted least-squares residuals to be equal to one for positive estimates. For the other two estimators, the difference between the commonly used test statistics and the newly proposed test statistics depends on the value of the mean sum of the weighted least-squares residuals.

In a simulation study we compared the commonly used tests and the newly proposed tests with regard to their actual significance level in the context of clinical trials which evaluated the efficacy of a vaccine for the prevention of tuberculosis already discussed in Berkey *et al.* [1]. The outcome measure was the logarithmic relative risk and, besides the 'usual' within-trial variance estimates of the estimated log relative risk, 'smoothed' within-trial variance estimates proposed by Berkey *et al.* [1] are considered. The simulation study showed that the newly proposed tests satisfactorily keep the nominal significance level irrespective of the choice of the between-trial variance estimator. The findings were confirmed in an extended simulation study where we varied the number of trials as well as the values of the covariate. In the setting of the example the use of the 'smoothed' within-trial estimates yielded better results with regard to the actual significance levels of the tests whereas in the extended simulation study we did not observe significant differences by using the 'usual' or the 'smoothed' within-trial variance estimates.

6.1. Meta-regression with more than one covariate

The proposed new standard error estimates for the regression parameter estimates can be easily extended to the case of more than one covariate in the meta-regression. The new standard error estimates can be seen as the product of the usual standard error estimate in the meta-regression with the mean sum of the weighted least-squares residuals, Thus, in the case of more than one covariate one has simply to calculate the weighted least-squares residuals along the lines given in the Appendix A2 to obtain the quadratic form q where now the matrix Z contains the different covariates and the trace of the matrix P_1 is equal to the number of trials minus the number of regression parameters. Moreover, the trace of P_1 yields the number of degrees of freedom for the t-distribution used in the corresponding confidence intervals and hypothesis tests, respectively.

6.2. Within-trial variance estimates for other outcome measures

In the simulation study, especially for the setting of the example, we have observed that the use of the 'smoothed' within-trial variance estimates improves the hypothesis tests with regard to their actual significance levels compared to the pre-chosen level when the logarithmic relative risk is used. The 'smoothed' within-trial variance estimates for the logarithmic relative risk proposed by Berkey *et al.* [1] reduce the correlation between the estimate of the logarithmic relative risk and its variance estimate in a trial. The relative weights of the trials induced by using the 'smoothed' within-trial variance estimates nearly correspond to relative weights induced by using the sample size as can been seen from Table V.

Berkey *et al.* [1] also proposed a 'smoothed' within-trial variance estimator for the logarithmic odds ratio. For the risk difference, an analogue 'smoothed' within-trial variance estimator

can be constructed as

$$\hat{\xi}_i^* = \frac{1}{n_{vi}} \frac{1}{k} \sum_{i=1}^{k} \frac{a_i}{n_{vi}} + \frac{1}{n_{ci}} \frac{1}{k} \sum_{i=1}^{k} \frac{c_i}{n_{ci}}$$

using the notation from Section 4. In both cases the correlation between the estimated treatment effect and its variance estimate is reduced.

When treatment effects as mean or mean difference for normal response are considered, the variance estimate is stochastically independent of the estimated treatment effect so that a 'smoothed' variant of a within-trial variance estimator does not have to be taken into account.

Comprehensively, by using the weighted least-squares approach in the random effects meta-regression, it is necessary to calculate the value of the mean sum of the weighted least-squares residuals to improve the confidence intervals and the hypothesis tests on the regression parameters and to use the t-distribution with degrees of freedom equal to the number of trials minus the number of regression parameters as the test distribution.

The performance of a meta-regression in practice raises some problems such as the appropriate choice of the covariates or the appropriate functional relation between the treatment effects and the covariates which should be modelled. The limitations and pitfalls in interpretation of meta-regression analyses are extensively discussed by Thompson and Higgins [15].

APPENDIX

Let A be a real-valued $(m \times n)$-matrix, then A' is the transpose of A and $\mathrm{rk}(A)$ the rank of A. For a quadratic matrix A let denote A^{-1} the inverse of A and $\mathrm{tr}(A)$ the trace of A.

A1. Method of moments estimator for the between-trial variance in the case of more than one covariate

Let us consider the general random effects meta-regression for meta-analysis

$$y \sim \mathrm{N}(X\gamma, \tau^2 I_k + \Delta)$$

with $y = (y_1, \ldots, y_k)'$, X the $(k \times r)$-dimensional regressor matrix with $\mathrm{rk}(X) = r < k - 1$, γ the unknown parameter vector of the fixed effects, τ^2 stands for the between-trial variance, I_k is the $(k \times k)$-dimensional identity matrix, and Δ is a $(k \times k)$-dimensional diagonal matrix with entries ξ_i, $i = 1, \ldots, k$, that is, Δ contains the within-trial variances.

The heterogeneity statistic can be expressed as a quadratic form in y and has the matrix representation

$$Q = y'P'\Delta^{-1}Py \quad \text{with } P = (I_k - X(X'\Delta^{-1}X)^{-1}X'\Delta^{-1})$$

The expected value of Q in the general meta-regression model (reference [16], p. 54) is given as

$$\begin{aligned} E(Q) &= \mathrm{tr}(P'\Delta^{-1}P\,\mathrm{cov}(y)) + \mathrm{tr}(P'\Delta^{-1}PX\gamma\gamma'X') \\ &= \mathrm{tr}((I_k - \Delta^{-1}X(X'\Delta^{-1}X)^{-1}X')\Delta^{-1}(I_k - X(X'\Delta^{-1}X)^{-1}X'\Delta^{-1})\,\mathrm{cov}(y)) \\ &= k - r + \tau^2 f(X, \Delta^{-1}) \end{aligned}$$

with $f(X, \Delta^{-1}) = \mathrm{tr}(\Delta^{-1}) - \mathrm{tr}((X'\Delta^{-1}X)^{-1}X'\Delta^{-2}X)$.

Consequently, the method of moments estimator is given as

$$\hat{\tau}^2 = \frac{Q - (k - r)}{f(X, \Delta^{-1})}$$

A2. Derivation of the variance estimators and the test distribution in model (1)

In matrix notation the random effects meta-regression with one covariate (1) can be written as

$$y \sim \mathrm{N}(Z\gamma, \Lambda)$$

with $y = (y_1, \ldots, y_k)'$, $Z = (1_k, x)$, $1_k = (1, \ldots, 1)'$, $x = (x_1, \ldots, x_k)'$, $\gamma = (\alpha, \beta)'$ and $\Lambda = \mathrm{diag}(\tau^2 + \xi_i)_{i=1,\ldots,k} = \mathrm{diag}(w_i^{-1})_{i=1,\ldots,k}$. The weighted least-squares estimators of α and β are given by

$$\gamma_{\mathrm{WLS}} = \begin{pmatrix} \alpha_{\mathrm{WLS}} \\ \beta_{\mathrm{WLS}} \end{pmatrix} = (Z'\Lambda^{-1}Z)^{-1}Z'\Lambda^{-1}y$$

and the variances and covariance of α_{WLS} and β_{WLS} are the corresponding components of the (2×2)-matrix $(Z'\Lambda^{-1}Z)^{-1}$. Let us consider the quadratic form from (12), that is

$$q = \frac{1}{k-2} \sum_{i=1}^{k} w_i (y_i - \alpha_{\mathrm{WLS}} - \beta_{\mathrm{WLS}} x_i)^2$$

This quadratic form can be written in matrix notation as

$$q = \frac{1}{k-2}\, y' P_1' \Lambda^{-1} P_1 y$$

with $P_1 = I_k - Z(Z'\Lambda^{-1}Z)^{-1}Z'\Lambda^{-1}$ and I_k the $(k \times k)$-identity matrix. Note that the matrix P_1 is idempotent, that is, $P_1^2 = P_1$.

Lemma 1
Under normality of y, the quadratic form q and the weighted least-squares estimator γ_{WLS} are stochastically independent.

Proof
It is sufficient to show (reference [16], p. 227) that the following equation holds:

$$P_1'\Lambda^{-1}P_1 \,\mathrm{cov}(y)\Lambda^{-1}Z(Z'\Lambda^{-1}Z)^{-1} = 0$$

Let us consider $P_1\Lambda\Lambda^{-1}Z = P_1 Z = 0$. Thus, the assertion follows.

Lemma 2
Under normality of y, the quadratic form $(k-2)q$ is χ^2-distributed with $k-2$ degrees of freedom.

Proof
Let us first consider

$$\Lambda P_1' \Lambda^{-1} = I_k - Z(Z'\Lambda^{-1}Z)^{-1}Z'\Lambda^{-1} = P_1$$

then it holds

$$(\Lambda P_1' \Lambda^{-1} P_1)^2 = (P_1 P_1)^2 = P_1 = \Lambda P_1' \Lambda^{-1} P_1$$

With $\mathrm{tr}(\Lambda P_1' \Lambda^{-1} P_1) = \mathrm{tr}(P_1) = k - 2$ and $P_1 Z = 0$, the assertion follows (reference [16], p. 197).

REFERENCES

1. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
3. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677.
4. Fleiss JL. The statistical basis of meta-analysis. *Statistical Methods in Medical Research* 1993; **2**:121–145.
5. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
6. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
7. Böckenhoff A, Hartung J. Some corrections of the significance level in meta-analysis. *Biometrical Journal* 1998; **8**:937–947.
8. Li Y, Shi L, Roth HD. The bias of the commonly-used estimate of variance in meta-analysis. *Communications in Statistics — Theory and Methods* 1994; **23**:1063–1085.
9. Ziegler S, Koch A, Victor N. Deficits and remedy of the standard random effects methods in meta-analysis. *Methods of Information in Medicine* 2001; **40**:148–155.
10. Hartung J, Knapp G. On tests of the overall treatment effect in the meta-analysis with normally distributed responses. *Statistics in Medicine* 2001; **20**:1771–1782.
11. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine* 2001; **20**:3875–3889.
12. DuMouchel WH, Harris JE. Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association* 1983; **78**:293–308.
13. Morris CN. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association* 1983; **78**:47–55.
14. Pettigrew HM, Gart JJ, Thomas DG. The bias and higher cumulants of the logarithm of a binomial variate. *Biometrika* 1986; **73**:425–435.
15. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002; **21**:1559–1573.
16. Mathai AM, Provost SB. *Quadratic Forms in Random Variables*. Marcel Dekker: New York, 1992.

