On the Meta-Analysis of Treatment Differences in Heteroscedastic Samples

[1]**Makambi, Kepher Henry**; [2]**Argaç, Doğan**; [3]**Hartung, Joachim**

[1]Department of Mathematics and Statistics,

Jomo Kenyatta University of Agriculture and Technology

P.O. Box 62000,Nairobi, Kenya.

[2,3]Department of Statistics, University of Dortmund,

D-44221 Dortmund, Germany

Summary

In many meta-analysis cases the estimator of the overall effect in independent trials or experiments leads to unjustified significant results. This paper considers trials with two arms where the summary statistic of interest is either the mean difference or the risk difference. By using convexity principles of the relevant composed functions and the moments of the chi-square distribution, corrections are made on the estimated standard deviation of the estimator of the overall treatment difference. It is shown, analytically and by simulations, that by making such corrections on the estimated standard deviation, significance levels are attained which are relatively closer to the nominal level.

Key Words: Mean difference; risk difference; convexity; significance levels

---

To appear in:

B I O M E T R I C A L L E T T E R S

1. Introduction and Notations

There are many areas (e.g. medicine, epidemiology and education) where the combination of results from different trials (studies or experiments) has become common. For example, a situation may arise where one has to assess the overall treatment difference when samples from the different trials are either homoscedastic or heteroscedastic. Recent studies by Li et al. (1994) and Boeckenhoff/Hartung (1998) attest to the fact that there is a systematic overestimate in the significance levels when combining studies in fixed effects models which may be due to the underestimate of the variance of the estimator of the overall treatment mean. If one considers trials which are comparative in nature, then measures of the common treatment difference may take different forms, for instance, mean differences or effect sizes for quantitative data, and risk differences, (logarithm) relative risks, or (logarithm) odds ratios for binary data. By considering the mean difference and the risk difference we show both analytically and by simulations that by making corrections on the estimated standard deviation of the overall mean difference (or overall risk difference) significance levels can be obtained which are relatively closer to the nominal level.

Suppose there are K "two-armed" (multicenter) trials in a meta-analysis.Let x_{jil} be the lth observation in arm j of trial i; $i = 1, \ldots, K$, $j = 1, 2$, $l = 1, \ldots, n_{ji}$ where n_{ji} is the total number of observations in arm j of trial i. Then the mean of the jth arm of study i is $\bar{x}_{ji} = \sum_{l=1}^{n_{ji}} x_{jil}/n_{ji} \sim (\mu_{ji}, \sigma^2_{\epsilon_{ji}}/n_{ji})$, and we will write $\sigma^2_{ji} = \sigma^2_{\epsilon_{ji}}/n_{ji}$. Define $y_i = \bar{x}_{1i} - \bar{x}_{2i} = y_{1i} - y_{2i}$, where $y_{ji} \sim (\mu_{ji}, \sigma^2_{ji})$, $i = 1, \ldots, K$, $j = 1, 2$, is one of the summary statistics available from arm j of study i for a meta-analysis.
Further, define

$$y_i = \mu + e_i, \;\; i = 1, \ldots, K, \tag{1}$$

where $e_i \sim (0, \sigma_i^2)$, and it is assumed that $\mu = \mu_{1i} - \mu_{2i}$ is common in all the studies. In this formulation, y_i could be the mean difference for quantitative data or the risk difference for binary data.

2. Estimation

Normal Data

Let $y_{ji} \sim \mathcal{N}(\mu_{ji}, \sigma_{ji}^2)$, $i = 1, \ldots, K$, $j = 1, 2$ ($\mu_{ji} = \mu_j$ independent of trial number i); so that $y_i \sim \mathcal{N}(\mu, \sigma_i^2)$, with $\sigma_i^2 = \sigma_{1i}^2 + \sigma_{2i}^2$; and $e_i \sim \mathcal{N}(0, \sigma_i^2)$, for $i = 1, \ldots, K$.

The best estimator of μ in each trial is the individual sample treatment difference

$$\hat{\mu}_i = \hat{\mu}_{1i} - \hat{\mu}_{2i} = y_{1i} - y_{2i};$$

Due to variation in sample sizes and precision of the trials, and absence of treatment-by-centre interaction, the best estimator of the underlying treatment difference (that is common to all trials) is a weighted estimate, namely

$$\hat{\mu} = \frac{1}{\sum_{i=1}^K 1/\sigma_i^2} \sum_{i=1}^K \frac{1}{\sigma_i^2} \cdot y_i, \tag{2}$$

with the associated variance (cf: Whitehead/Whitehead, 1991)

$$\sigma_{\hat{\mu}}^2 = \frac{1}{\sum_{i=1}^K 1/\sigma_i^2} \tag{3}$$

Let $s_{ji}^2 = s_{\epsilon_{ji}}^2 / n_{ji}$, with $s_{\epsilon_{ji}}^2 = (1/n_{ji} - 1) \cdot \sum_{l=1}^{n_{ji}} (x_{jil} - \bar{x}_{ji})^2$, be the estimate of σ_{ji}^2, for $i = 1, \ldots, K$, $j = 1, 2$. Then the estimate of σ_i^2 is $\hat{\sigma}_i^2 = s_i^2 = s_{1i}^2 + s_{2i}^2$. Therefore, the estimate of μ is given by

$$\hat{\mu}^* = \frac{1}{\sum_{i=1}^K 1/\hat{\sigma}_i^2} \sum_{i=1}^K \frac{1}{\hat{\sigma}_i^2} \cdot y_i, \tag{4}$$

and

$$\hat{\sigma}_{\hat{\mu}}^2 = \frac{1}{\sum_{i=1}^{K} 1/\hat{\sigma}_i^2}, \tag{5}$$

Further, we have that

$$Var(\hat{\sigma}_i^2) = Var(\hat{\sigma}_{1i}^2) + Var(\hat{\sigma}_{2i}^2) = \frac{2}{n_{1i}-1} \cdot \sigma_{1i}^4 + \frac{2}{n_{2i}-1} \cdot \sigma_{2i}^4 \tag{6}$$

which is estmated by

$$\widehat{Var}(\hat{\sigma}_i^2) = \widehat{Var}(\hat{\sigma}_{1i}^2) + \widehat{Var}(\hat{\sigma}_{2i}^2) = \frac{2}{n_{1i}-1} \cdot s_{1i}^4 + \frac{2}{n_{2i}-1} \cdot s_{2i}^4 \tag{7}$$

Binomial Data

For binary data, let y_{ji} be binomially distributed with parameters n_{ji} and p_{ji}, $i = 1, \ldots, K$, j=1,2. Therefore, $y_i = \hat{\mu}_{1i} - \hat{\mu}_{2i} = \hat{p}_{1i} - \hat{p}_{2i} \overset{approx}{\sim} \mathcal{N}(\mu, \sigma_i^2)$, with $\mu = p_{1i} - p_{2i} = p_1 - p_2$, assumed identical in all the trials, i=1,…,K and $\hat{p}_{ji} = y_{ji}/n_{ji}$. Here

$$\hat{\mu}^* = \frac{1}{\sum_{i=1}^{K} 1/\hat{\sigma}_i^2} \sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2} \cdot y_i, \tag{8}$$

with

$$\hat{\sigma}_i^2 = \hat{\sigma}_{1i}^2 + \hat{\sigma}_{2i}^2 = \frac{1}{n_{1i}-1} \cdot (\hat{p}_{1i} - \hat{p}_{1i}^2) + \frac{1}{n_{2i}-1} \cdot (\hat{p}_{2i} - \hat{p}_{2i}^2). \tag{9}$$

It is sufficient in our case to approximate $Var(\hat{\sigma}_{ji}^2)$ by the delta-method, thus

$$Var(\hat{\sigma}_{ji}^2) \approx \left(\frac{\partial \hat{\sigma}_{ji}^2}{\partial \hat{p}_{ji}} |_{\hat{p}_{ji}=p_{ji}} \right)^2 \cdot \sigma_{ji}^2 = \left(\frac{1-2p_{ji}}{n_{ji}-1} \right)^2 \cdot \frac{1}{n_{ji}} p_{ji}(1-p_{ji}), \tag{10}$$

which is estimated by replacing p_{ji} with $\hat{p}_{ji}$.

In both the normal and binomial populations, to set confidence intervals and testing hypotheses, we use the corresponding statistic

$$T = \frac{\hat{\mu}^*}{\hat{\sigma}_{\hat{\mu}}} \overset{approx}{\sim} \mathcal{N}(\mu, 1). \tag{11}$$

The estimator $\hat{\sigma}_{\hat{\mu}}$ is biased and underestimates $\sigma_{\hat{\mu}}$. This can easily be shown by using the concavity of $\hat{\sigma}_{\hat{\mu}}^2$ in $\hat{\sigma}_i^2$ and Jensen's inequality (cf: Hartung, 1977 and Li, et al.,1994). That is,

$$E(\hat{\sigma}_{\hat{\mu}}) \leq \sigma_{\hat{\mu}}.$$

Trials where the estimator $\hat{\sigma}_{\hat{\mu}}$, which underestimates $\sigma_{\hat{\mu}}$, is used in obtaining T are often bound to be unjustifiably significant. Tables 1 and 2 below give some actual simulated significance levels for testing the hypothesis $H_0 : \mu = 0$ against a two sided alternative $H_1 : \mu \neq 0$ at $\alpha = 0.05$ for different constellations of (n_{1i}, n_{2i}) and $(\sigma^2_{\epsilon_{1i}}, \sigma^2_{\epsilon_{2i}})$; $i = 1, \ldots, K$, for K=3 for normal and binomial data.

Table 1: Actual simulated significance levels for K=3 (Normal data).

(n_{11}, n_{21})	(n_{12}, n_{22})	(n_{13}, n_{23})	$(\sigma^2_{\epsilon_{11}}, \sigma^2_{\epsilon_{21}})$	$(\sigma^2_{\epsilon_{12}}, \sigma^2_{\epsilon_{22}})$	$(\sigma^2_{\epsilon_{13}}, \sigma^2_{\epsilon_{23}})$	$\hat{\alpha}\%$
(5,6)	(6,7)	(7,5)	(1,4)	(3,4)	(5,4)	10.3
			(1,5)	(3,3)	(5,1)	10.5
			(1,10)	(3,30)	(5,50)	13.1
(10,10)	(10,10)	(10,10)	(1,4)	(3,4)	(5,4)	8.3
			(1,5)	(3,3)	(5,1)	8.1
			(1,10)	(3,30)	(5,50)	9.2

Table 2: Actual simulated significance levels for K=3 (Binomial data).

(n_{11}, n_{21})	(n_{12}, n_{22})	(n_{13}, n_{23})	(p_1, p_2)	$\hat{\alpha}\%$
(7,13)	(10,7)	(15,10)	(0.3,0.3)	6.0
			(0.4,0.4)	6.8
			(0.5,0.5)	7.1
			(0.6,0.6)	6.6
			(0.7,0.7)	6.2

All of the attained significance levels given in Tables 1 and 2 are larger than the expected nominal level of 5%. Our concern is in the methods which will make the attained significance levels closer to the nominal level.

3. Some Theoretical Results

Define on $I\!R_+^K$ the function $f(x) = \sum_{i=1}^K 1/x_i$, then f is convex, and $h(x) = 1/f(x)$ is quasi-concave. Next, define $h(0) = 0$, then $h(\lambda x) = \lambda \cdot h(x)$, $\lambda \geq 0$, $x > 0$, implies that h is positively homogeneous; so together with the quasi-concavity it follows that h is concave (cf: Hartung, 1976, section1).

By Jensen's inequality, if f is convex, then $Ef(x) \geq f(Ex)$ and the reverse is true if f is concave.

Now, consider $\hat{\sigma}_i^2 = \hat{\sigma}_{1i}^2 + \hat{\sigma}_{2i}^2$ as given early. By Patnaik (1949), the statistic $\nu_i \hat{\sigma}_i^2 / E(\hat{\sigma}_i^2) \sim \chi^2_{\nu_i}$, where

$$\nu_i = 2 \cdot \frac{\left(E(\hat{\sigma}_{1i}^2) + E(\hat{\sigma}_{2i}^2)\right)^2}{Var(\hat{\sigma}_{1i}^2) + Var(\hat{\sigma}_{2i}^2)};$$

which is estimated by

$$\hat{\nu}_i = 2 \cdot \frac{\left(\hat{\sigma}_{1i}^2 + \hat{\sigma}_{2i}^2\right)^2}{\widehat{Var}(\hat{\sigma}_{1i}^2) + \widehat{Var}(\hat{\sigma}_{2i}^2)};$$

for $i = 1, \ldots, K$.

This facilitates the definition of the following approximate moments and inverse moments of the chi-square distribution(cf: Patel et al., 1976):

$$\begin{aligned} E\left(\hat{\sigma}_i\right) &= \gamma_{\hat{\nu}_i}^{-1} \cdot \sigma_i; \quad \gamma_{\hat{\nu}_i} = \sqrt{\frac{\hat{\nu}_i}{2}} \cdot \frac{\Gamma(\hat{\nu}_i/2)}{\Gamma(\hat{\nu}_i/2 + 1/2)} && (12) \\ E\left(\hat{\sigma}_i^4\right) &= b_{\hat{\nu}_i+1} \cdot \sigma_i^4; \quad b_{\hat{\nu}_i+1} = \frac{\hat{\nu}_i + 2}{\hat{\nu}_i} && (13) \\ E\left(\hat{\sigma}_i^{-2}\right) &= c_{\hat{\nu}_i+1} \cdot \frac{1}{\sigma_i^2}; \quad c_{\hat{\nu}_i+1} = \frac{\hat{\nu}_i}{\hat{\nu}_i - 2}, && (14) \end{aligned}$$

Now, with $\hat{\sigma}_{\hat{\mu}}^2 = \left(\sum_{i=1}^K 1/\hat{\sigma}_i^2\right)^{-1}$, we have

$$Var(\hat{\sigma}_{\hat{\mu}}^2) \leq \left(\sum_{i=1}^K \sqrt{\frac{\hat{\nu}_i}{\hat{\nu}_i + 2}} \cdot \frac{1}{\sigma_i^2}\right)^{-2} - \left(\sum_{i=1}^K \frac{\hat{\nu}_i}{\hat{\nu}_i - 2} \cdot \frac{1}{\sigma_i^2}\right)^{-2} \qquad (15)$$

This can be proved as follows:

$$Var(\hat{\sigma}_{\hat{\mu}}^2) = E\left(\left(\sum_{i=1}^K \frac{1}{\hat{\sigma}_i^2}\right)^{-1}\right)^2 - \left(E\left(\sum_{i=1}^K \frac{1}{\hat{\sigma}_i^2}\right)^{-1}\right)^2$$

$$\leq \ E\left(\left(\left(\sum_{i=1}^{K}\frac{1}{\hat{\sigma}_i^2}\right)^{-1}\right)^2\right) - \left(\sum_{i=1}^{K}E\left(\frac{1}{\hat{\sigma}_i^2}\right)\right)^{-2}$$

by using the convexity of $\left(\sum_{i=1}^{K}1/\hat{\sigma}_i^2\right)^{-1}$ in $\hat{\sigma}_i^{-2}$. Using first (13) and then (14), we have

$$\begin{aligned}
Var(\hat{\sigma}_{\hat{\mu}}^2) &\leq E\left(\sum_{i=1}^{K}\frac{1}{\sqrt{\hat{\sigma}_i^4}}\right)^{-2} - \left(\sum_{i=1}^{K}c_{\hat{\nu}_i+1}\cdot\frac{1}{\sigma_i^2}\right)^{-2} \\
&\leq \left(\sum_{i=1}^{K}\frac{1}{\sqrt{b_{\hat{\nu}_i+1}\cdot\sigma_i^4}}\right)^{-2} - \left(\sum_{i=1}^{K}c_{\hat{\nu}_i+1}\cdot\frac{1}{\sigma_i^2}\right)^{-2} \\
&= \left(\sum_{i=1}^{K}\sqrt{\frac{\hat{\nu}_i}{\hat{\nu}_i+2}}\cdot\frac{1}{\sigma_i^2}\right)^{-2} - \left(\sum_{i=1}^{K}\frac{\hat{\nu}_i}{\hat{\nu}_i-2}\cdot\frac{1}{\sigma_i^2}\right)^{-2} \qquad (16)
\end{aligned}$$

by the concavity of $\left(\hat{\sigma}_{\hat{\mu}}^2\right)^2$ in $\hat{\sigma}_i^4$; which is seen to be similar to $h(x)$ above.

Now, let

$$\hat{\phi} = \left(\sum_{i=1}^{K}\sqrt{\frac{\hat{\nu}_i}{\hat{\nu}_i+2}}\cdot\frac{1}{\hat{\sigma}_i^2}\right)^{-2} - \left(\sum_{i=1}^{K}\frac{\hat{\nu}_i}{\hat{\nu}_i-2}\cdot\frac{1}{\hat{\sigma}_i^2}\right)^{-2} \qquad (17)$$

with $\phi = \left(\sum_{i=1}^{K}\sqrt{\hat{\nu}_i/(\hat{\nu}_i+2)}\cdot 1/\sigma_i^2\right)^{-2} - \left(\sum_{i=1}^{K}\hat{\nu}_i/(\hat{\nu}_i-2)\cdot 1/\sigma_i^2\right)^{-2}$ as in (14). Define a class of estimators of $\sigma_{\hat{\mu}}$ by

$$\hat{\sigma}_{\hat{\mu}}(\hat{\phi},\tau) = \sqrt{\left(\hat{\sigma}_{\hat{\mu}}^2 + \tau\cdot\sqrt{\hat{\phi}}\right)}, \qquad (18)$$

with $\hat{\phi}$ defined in (17) and the control parameter $\tau > 0$. It is clear that $\hat{\sigma}_{\hat{\mu}}(\hat{\phi},\tau) \geq \hat{\sigma}_{\hat{\mu}}$.

Further, consider the following results:

$$\begin{aligned}
\text{i)} \quad E(\hat{\sigma}_{\hat{\mu}}) &= E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{\hat{\sigma}_i^2}\right)^{-1}} \\
&\leq E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{\gamma_{\hat{\nu}_i}^2\cdot\hat{\sigma}_i^2}\right)^{-1}} \\
&\leq \sqrt{\left(\sum_{i=1}^{K}\frac{1}{\gamma_{\hat{\nu}_i}^2\cdot(E\sqrt{\hat{\sigma}_i^2})^2}\right)^{-1}}\ ; \ \text{since } \hat{\sigma}_{\hat{\mu}} \text{ is concave in } \hat{\sigma}_i
\end{aligned}$$

$$= \quad \sigma_{\hat{\mu}}$$

$$\text{ii)} \quad \sigma_{\hat{\mu}} = \sqrt{\left(\sum_{i=1}^{K}\frac{1}{\sigma_i^2}\right)^{-1}}$$

$$= \sqrt{\left(\sum_{i=1}^{K}\frac{1}{c_{\hat{\nu}_i+1}}\cdot E\left(\frac{1}{\hat{\sigma}_i^2}\right)\right)^{-1}}$$

$$\leq E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{c_{\hat{\nu}_i+1}}\cdot\frac{1}{\hat{\sigma}_i^2}\right)^{-1}}, \tag{19}$$

due to the convexity of $\hat{\sigma}_{\hat{\mu}}$ in $1/\hat{\sigma}_i^2$. From (i) and (ii) above, we have the following

$$E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{\hat{\sigma}_i^2}\right)^{-1}} \leq \sigma_{\hat{\mu}} \leq E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{c_{\hat{\nu}_i+1}}\cdot\frac{1}{\hat{\sigma}_i^2}\right)^{-1}} \leq E\sqrt{\left(\sum_{i=1}^{K}\frac{1}{c_{\hat{\nu}_i+1}\cdot\gamma_{\hat{\nu}_i}^3}\cdot\frac{1}{\hat{\sigma}_i^2}\right)^{-1}} \tag{20}$$

4. Simulation Results

To demonstrate how the proposed methods perform, a simulation study is carried out with the number of trials, K=3, 6 and 9 for both the normal and binomial cases. Different constellations of unbalanced heteroscedastic samples are considered as shown in Tables 3.a., 3.b., 3.c. for the normal case and Tables 4.a., 4.b., 4.c. for the binomial case.

To get an impression of how these procedures perform for relatively large trials, we started with K=3 and made independent replications to give K=6, denoted by $2\times\omega$, and K=9, denoted by $3\times\omega$; see the Tables below. Further, for K=6, for example, replication was done such that $n_{11}=n_{14}$, $n_{21}=n_{24}$, $n_{12}=n_{15}$, $n_{22}=n_{25}$, $n_{13}=n_{16}$, $n_{23}=n_{26}$ and similarly for variances in the normal case. For K=9, $n_{11}=n_{14}=n_{17}$, $n_{21}=n_{24}=n_{27}$, $n_{12}=n_{15}=n_{18}$, $n_{22}=n_{25}=n_{28}$, $n_{13}=n_{16}=n_{19}$, $n_{23}=n_{26}=n_{29}$.

Used also in the Tables are the following representations:

$$\hat{\sigma}_{\hat{\mu}}(c) = \sqrt{\left(\sum_{i=1}^{K} \frac{1}{c_{\hat{\nu}_i+1}} \cdot \frac{1}{\hat{\sigma}_i^2}\right)^{-1}}, \quad \hat{\sigma}_{\hat{\mu}}(c\gamma^3) = \sqrt{\left(\sum_{i=1}^{K} \frac{1}{c_{\hat{\nu}_i+1} \cdot \gamma_{\hat{\nu}_i}^3} \cdot \frac{1}{\hat{\sigma}_i^2}\right)^{-1}}$$

and $\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, \tau)$ as in (18).

Table 3.a.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=3 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ with test statistic as in (11) and different estimators for the standard deviation with $a_1 = (n_{11}, n_{21})$, $a_2 = (n_{12}, n_{22})$, $a_3 = (n_{13}, n_{23})$ and $b_1 = (\sigma^2_{\epsilon_{11}}, \sigma^2_{\epsilon_{21}})$, $b_2 = (\sigma^2_{\epsilon_{12}}, \sigma^2_{\epsilon_{22}})$, $b_3 = (\sigma^2_{\epsilon_{13}}, \sigma^2_{\epsilon_{23}})$.

Test statistic, $T = \hat{\mu}^*/\Delta$										
Sample Sizes			Variances			$\hat{\alpha}\%$				
						Δ				
a_1	a_2	a_3	b_1	b_2	b_3	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(6,13)	(10,6)	(15,10)	(1,4)	(3,6)	(5,3)	6.0	7.9	6.0	4.6	5.4
			(10,40)	(30,60)	(50,30)	6.0	8.0	6.0	4.7	5.4
			(20,80)	(60,120)	(100,60)	6.1	8.3	6.6	5.0	5.8
(12,26)	(20,12)	(30,20)	(1,4)	(3,6)	(5,3)	5.7	6.6	5.7	4.4	5.4
			(10,40)	(30,60)	(50,30)	5.3	6.3	5.3	3.8	4.9
			(20,80)	(60,120)	(100,60)	5.8	6.6	5.8	4.4	5.5

From Tables 3.$a-c$ and 4.$a-c$, we see that results with $\hat{\sigma}_{\hat{\mu}}$ always overestimate the nominal significance level. This overestimate is relatively more pronounced for the normal case (Tables 3.$a - c$).

Using $\hat{\sigma}_{\hat{\mu}}(c)$ results in significance levels which are in the same order of magnitude with the levels of $\sigma_{\hat{\mu}}$; notice the large number of levels which are actually equal for the normal case, Tables 3.$a - c$. The results of $\hat{\sigma}_{\hat{\mu}}(c)$ in the binomial case are in the same order of magnitude with those of $\hat{\sigma}_{\hat{\mu}}$, Tables 4.$a - c$.

By using $\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$ and $\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$ we obtain further improvements of significance levels. The advantage with $\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$ is that we can obtain more improvements by varying the control parameter, τ.

There does not seem to be any sensitivity of the test statistics to changes in the number of trials and the error variances.

Table 3.b.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=6 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ with test statistic as in (11) and different estimators for the standard deviation.

Test statistic, $T = \hat{\mu}^*/\Delta$										
Sample Sizes			Variances			$\hat{\alpha}\%$				
						Δ				
$2\times a_1$	$2\times a_2$	$2\times a_3$	$2\times b_1$	$2\times b_2$	$2\times b_3$	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(6,13)	(10,6)	(15,10)	(1,4)	(3,6)	(5,3)	6.8	9.0	6.8	5.3	6.3
			(10,40)	(30,60)	(50,30)	6.7	8.8	6.6	5.1	6.0
			(20,80)	(60,120)	(100,60)	6.2	8.0	6.1	4.8	5.5
(12,26)	(20,12)	(30,20)	(1,4)	(3,6)	(5,3)	5.8	6.7	5.8	4.4	5.5
			(10,40)	(30,60)	(50,30)	5.7	7.1	5.8	4.4	5.5
			(20,80)	(60,120)	(100,60)	5.4	6.5	5.5	4.0	5.2

Table 3.c.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=9 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ with test statistic as in (11) and different estimators for the standard deviation.

Test statistic, $T = \hat{\mu}^*/\Delta$										
Sample Sizes			Variances			$\hat{\alpha}\%$				
						Δ				
$3\times a_1$	$3\times a_2$	$3\times a_3$	$3\times b_1$	$3\times b_2$	$3\times b_3$	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(6,13)	(10,6)	(15,10)	(1,4)	(3,6)	(5,3)	6.7	8.7	6.7	5.1	6.0
			(10,40)	(30,60)	(50,30)	6.7	9.1	6.8	5.2	6.1
			(20,80)	(60,120)	(100,60)	6.6	8.6	6.5	5.1	6.0
(12,26)	(20,12)	(30,20)	(1,4)	(3,6)	(5,3)	5.8	6.7	5.7	4.3	5.4
			(10,40)	(30,60)	(50,30)	5.8	6.7	5.7	4.3	5.4
			(20,80)	(60,120)	(100,60)	5.9	6.9	5.9	4.5	5.6

Table 4.a.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=3 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ with test statistic like (11) and different estimators for the standard deviation with $a_1 = (n_{11}, n_{21})$, $a_2 = (n_{12}, n_{22})$, $a_3 = (n_{13}, n_{23})$.

Test statistic, $T = \hat{\mu}^*/\Delta$								
Sample Sizes				$\hat{\alpha}\%$				
				Δ				
a_1	a_2	a_3	(p_1, p_2)	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(7,13)	(10,7)	(15,10)	(0.3,0.3)	5.1	6.0	5.8	5.0	4.7
			(0.4,0.4)	5.9	6.8	6.6	6.0	5.7
			(0.5,0.5)	6.2	7.1	7.0	6.5	6.4
			(0.6,0.6)	5.7	6.7	6.5	5.9	5.8
			(0.7,0.7)	5.2	6.2	6.0	5.1	4.7
(15,25)	(20,15)	(30,20)	(0.3,0.3)	4.9	5.4	5.3	4.9	4.8
			(0.4,0.4)	5.2	5.6	5.6	5.5	5.5
			(0.5,0.5)	5.7	6.0	6.0	5.9	5.9
			(0.6,0.6)	5.2	5.6	5.6	5.5	5.5
			(0.7,0.7)	5.2	5.6	5.5	5.2	5.1

Table 4.b.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=6 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$. with test statistic like (11) and different estimators for the standard deviation.

Test statistic, $T = \hat{\mu}^*/\Delta$								
Sample Sizes				$\hat{\alpha}\%$				
				Δ				
$2 \times a_1$	$2 \times a_2$	$2 \times a_3$	(p_1, p_2)	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(7,13)	(10,7)	(15,10)	(0.3,0.3)	6.1	6.6	6.5	5.5	5.2
			(0.4,0.4)	6.3	6.9	6.8	6.3	6.2
			(0.5,0.5)	6.7	7.2	7.2	6.8	6.7
			(0.6,0.6)	6.5	7.0	6.9	6.4	6.3
			(0.7,0.7)	5.7	6.3	6.1	5.1	4.9
(15,25)	(20,15)	(30,20)	(0.3,0.3)	5.2	5.6	5.6	5.2	5.2
			(0.4,0.4)	5.4	5.6	5.6	5.4	5.4
			(0.5,0.5)	5.5	5.8	5.8	5.7	5.7
			(0.6,0.6)	5.8	6.1	6.1	5.9	5.9
			(0.7,0.7)	5.2	5.7	5.6	5.2	5.1

Table 4.c.: Simulated actual significance levels (10 000 runs) at nominal level $\alpha = 5\%$ for K=9 and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ with test statistic like (11) and different estimators for the standard deviation.

Test statistic, $T = \hat{\mu}^*/\Delta$								
Sample Sizes				$\hat{\alpha}\%$				
				Δ				
$3 \times a_1$	$3 \times a_2$	$3 \times a_3$	(p_1, p_2)	$\sigma_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\mu}}(c)$	$\hat{\sigma}_{\hat{\mu}}(\hat{\phi}, 0.5)$	$\hat{\sigma}_{\hat{\mu}}(c\gamma^3)$
(7,13)	(10,7)	(15,10)	(0.3,0.3)	6.0	6.6	6.4	5.2	5.0
			(0.4,0.4)	6.9	7.6	7.6	6.8	6.8
			(0.5,0.5)	6.8	7.3	7.3	6.8	6.7
			(0.6,0.6)	6.6	7.1	7.1	6.5	6.4
			(0.7,0.7)	5.7	6.2	5.9	5.1	4.9
(15,25)	(20,15)	(30,20)	(0.3,0.3)	5.5	5.7	5.7	5.4	5.4
			(0.4,0.4)	6.0	6.3	6.3	6.0	6.0
			(0.5,0.5)	6.3	6.5	6.5	6.4	6.4
			(0.6,0.6)	5.4	5.6	5.6	5.4	5.4
			(0.7,0.7)	5.4	5.8	5.7	5.4	5.3

5. Conclusion

In this article we have illustrated analytically and by simulations that attained significance levels could be improved by using suitable weights for the estimated standard deviation of the estimator of the overall treatment difference. The use of the methods developed is recommended especially when the number of trials is small.

A further investigation in this direction is to find out which methods are suitable when the measure of treatment effect is, for example, the effect sizes. The extension of these procedures to cases when there is treatment-by-center-interaction is also possible.

References

Boeckenhoff, A.; Hartung, J. (1998): Some Corrections of the Significance Level in Meta-Analysis. *Biometrical Journal* 40, 937-947.

Hartung, J. (1976): Some Inequalities for a Random Function. *Teorija Verojatnostei i ee Primenenja* 21, 661-665.

Li, Y.; Shi, L.; Roth, H. D. (1994): The bias of commonly-used estimates of variance in meta-analysis. *Commun. Statistics- Theory Methods* 23,1063-1085.

Patel, J. K.; Kapadia, C. P.; Owen, D. B. (1976): *Handbook of Statistical Distributions.* Marcel Dekker, New York.

Patnaik, P. B. (1949): The non-central χ^2- and F-distributions and their applications. *Biometrika* 36,202-232.

Whitehead, A.; Whitehead, J. (1991): A general parametric approach to meta-analysis of randomised clinical trials. *Statist. Med.* 10, 1665-1677.

Acknowledgement: The authors[1,2] would like to thank the German Academic Exchange Programme (DAAD) and the Deutsche Forschungsgemeinschaft (SFB 475) for financial assistance.