Discussiones Mathematicae
Probability and Statistics 23 (2003) 123–145

ON SMALL SAMPLE INFERENCE FOR COMMON MEAN IN HETEROSCEDASTIC ONE-WAY MODEL

Viktor Witkovský, Alexander Savin

Institute of Measurement Science, Slovak Academy of Sciences
Dúbravská cesta 9, 841 04 Bratislava, Slovakia

e-mail: witkovsky@savba.sk
e-mail: savin@savba.sk

AND

Gejza Wimmer

Matej Bel University,
Tajovského 40, 974 01 Banská Bystrica, Slovakia
and
Mathematical Institute, Slovak Academy of Sciences
Štefánikova 49, 814 73 Bratislava, Slovakia

e-mail: wimmer@mat.savba.sk

Abstract

In this paper we consider and compare several approximate methods for making small-sample statistical inference on the common mean in the heteroscedastic one-way random effects model. The topic of the paper was motivated by the problem of interlaboratory comparisons and is also known as the (traditional) common mean problem. It is also closely related to the problem of multicenter clinical trials and meta-analysis. Based on our simulation study we suggest to use the approach proposed by Kenward & Roger (1997) as an optimal choice for construction of the interval estimates of the common mean in the heteroscedastic one-way model.

Keywords: interlaboratory trials, common mean, generalized p-values, Kenward-Roger method.

2000 Mathematics Subject Classification: 62F25, 62E15.

1. Introduction

In many applications the researcher has to make inference on the common overall effect (i.e. the common mean or the consensus mean) based on several experiments or samples. The topic of this paper was motivated by the problem of interlaboratory comparisons and is also known as the (traditional) common mean problem.

Consider that the measurements on virtually the same object of interest are made by $k \geq 2$ laboratories. The i-th laboratory repeats its measurements n_i times, $n_i \geq 2$. The laboratories may exhibit the between laboratory variability, as well as possibly heterogeneous within-laboratory variances (i.e. the case of heteroscedasticity of measurement errors). In this paper we will assume that the measurements follow normal distribution.

The results of typical interlaboratory studies are presented in Table 1 and Table 2, for more details see also Eberhardt, Reeve & Spiegelman (1989) and Rukhin & Vangel (1998).

Table 1. Selenium in non-fat milk powder data.

Method	$\bar{y}_i$	s_i^2	n_i	Method	$\bar{y}_i$	s_i^2	n_i
A	105.00	85.711	8	C	109.50	2.729	14
B	109.75	20.748	12	D	113.25	33.640	8

Table 2. Arsenic in oyster tissue data.

Lab	$\bar{y}_i$	$\sqrt{s_i^2}$	n_i	Lab	$\bar{y}_i$	$\sqrt{s_i^2}$	n_i	Lab	$\bar{y}_i$	$\sqrt{s_i^2}$	n_i
1	9.78	0.30	5	11	13.08	0.43	5	21	13.94	0.15	5
2	10.18	0.46	5	12	13.30	0.16	5	22	13.98	0.80	5
3	10.35	0.04	2	13	13.46	0.21	5	23	14.22	0.88	5
4	11.60	0.78	5	14	13.48	0.41	5	24	14.60	0.43	5
5	12.01	2.62	5	15	13.48	0.47	5	25	14.68	0.33	5
6	12.26	0.83	5	16	13.55	0.06	5	26	15.00	0.71	5
7	12.88	0.59	5	17	13.61	0.36	5	27	15.08	0.18	5
8	12.88	0.29	5	18	13.78	0.61	5	28	15.48	1.64	5
9	12.96	0.52	5	19	13.82	0.33	5				
10	13.00	0.86	5	20	13.86	0.28	5				

For analysis of such data we will consider here the one-way random effects model, which, in general, can be heteroscedastic and unbalanced:

$$Y_{ij} = \mu + b_i + \varepsilon_{ij}. \tag{1}$$

Here μ is the true, however unknown, measured value of the quantity of interest – the common mean, b_i are the random laboratory effects (biases) and $\varepsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, are mutually independent errors. The result of the (i, j)-th measurement is y_{ij} – the realization of the random variable Y_{ij}. We will also use the following notation:

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{and} \quad S_i^2 = \frac{1}{(n_i - 1)} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2, \tag{2}$$

the sample laboratory means and the sample laboratory variances, and their realized values $\bar{y}_i$ and s_i^2, $i = 1, \ldots, k$.

In the setup of the traditional common mean problem, it is assumed that $b_i \sim N(0, \sigma_B^2)$ are mutually independent and independent with all ε_{ij}. The variance components σ_i^2, $i = 1, \ldots, k$, and σ_B^2 are the nuisance parameters – the within laboratory and the between laboratory variances, respectively.

In particular situation the researcher can assume that the between laboratory variability can be neglected, that is $\sigma_B^2 = 0$, i.e.

$$Y_{ij} = \mu + \varepsilon_{ij}, \tag{3}$$

with $\varepsilon_{ij} \sim N(0, \sigma_i^2)$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$.

One possibility to check that the between laboratory variance can be neglected is to test the null hypothesis $H_0 : \sigma_B^2 = 0$ against the one-sided alternative $H_1 : \sigma_B^2 > 0$ in the model (1), or to construct the interval estimate for σ_B^2.

If the variance components σ_B^2 and σ_i^2 were known, the optimal estimator for the unknown common mean μ would be the generalized least squares estimator (GLS estimator) which is under the given assumptions MVUE — minimum variance unbiased estimator, see e.g. Searle, Casella & McCulloch (1992), that is

$$\hat{\mu}^{(GLS)} = \frac{\sum_{i=1}^{k} w_i \bar{Y}_i}{\sum_{i=1}^{k} w_i}. \tag{4}$$

Here $w_i = 1/Var(\bar{Y}_i)$ with $Var(\bar{Y}_i) = \sigma_B^2 + \sigma_i^2/n_i$, i.e. $\hat{\mu}^{(GLS)}$, the optimal estimator of μ, is the weighted average of k laboratory average values $\bar{Y}_i$ with the weights w_i inversely proportional to the variances of the individual laboratory averages. Under given assumptions the exact distribution of the estimator is known:

$$\hat{\mu}^{(GLS)} \sim N(\mu, \Phi), \tag{5}$$

where $\Phi = 1/w_\Sigma$ and $w_\Sigma = \sum_{i=1}^{k} w_i$. From that the standard statistical inference on μ could be performed.

If the variance components are unknown, the situation becomes more complicated. The natural idea, frequently used by researchers, is to estimate the unknown variance components σ_B^2 and σ_i^2, then to plug-in those estimates into (4) and get the two-stage estimate $\hat{\hat{\mu}}$ of the common mean μ. Natural estimators of variance components which could be considered are the ML or REML estimators, or other type of estimators, which are computationally more simple than the MLEs (e.g. Mandel-Paule estimators, MINQUEs or other quadratic estimators).

In order to construct an approximate interval estimate the researcher is typically assuming that approximately the following distributional property holds true:

$$\hat{\hat{\mu}} \overset{\text{appr.}}{\sim} N(\mu, \hat{\Phi}), \tag{6}$$

where $\hat{\Phi} = 1/\sum_{i=1}^{k} \hat{w}_i$, $\hat{w}_i = 1/(\hat{\sigma}_B^2 + \hat{\sigma}_i^2/n_i)$, and $\hat{\sigma}_B^2$ and $\hat{\sigma}_i^2$ are (consistent) estimators of the variance components σ_B^2 and σ_i^2. Then the approximate $(1-\alpha)\times 100\%$ confidence interval for the common mean μ is given by

$$\hat{\hat{\mu}} \pm z(1-\alpha/2)\sqrt{\hat{\Phi}}, \tag{7}$$

where $z(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of standard normal distribution. It was proved by several simulation studies, that the variance of the two-stage (or plug-in) estimator $\hat{\hat{\mu}}$ is underestimated by $\hat{\Phi}$, and as such, the interval estimate based on (7) leads to too narrow confidence intervals for the parameter of interest — the common mean μ.

In Figure 1 the empirical coverage probabilities of the nominal 95% interval estimator (7), calculated with the Mandel-Paule estimates of variance components, are presented for different situations. For more details on the design of the simulation study see the Appendix A. As expected, in all considered situations the empirical coverage probabilities are bellow the nominal 95% level.

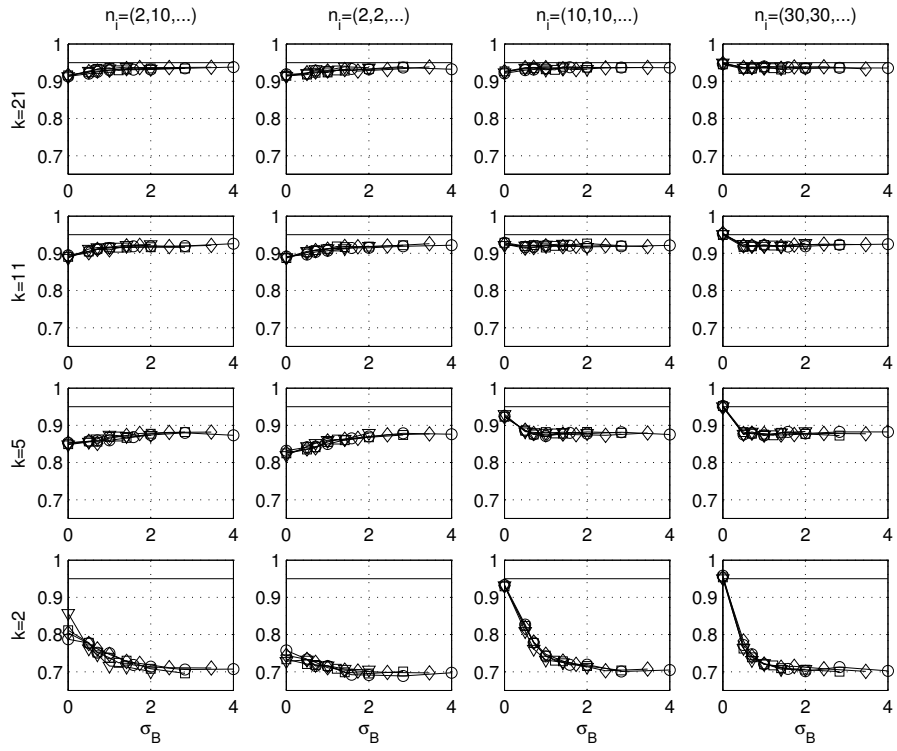

Figure 1. The empirical coverage probabilities of the 95% confidence intervals (7) calculated with the Mandel-Paule estimates of variance components, based on 10,000 Monte Carlo runs for each specific design as specified in the Appendix A. Here we use the symbol $\bigtriangledown$ for designs with $\sigma_k^2 = 1$, $\square$ for designs with $\sigma_k^2 = 2$, $\diamond$ for designs with $\sigma_k^2 = 3$, and $\bigcirc$ for designs with $\sigma_k^2 = 4$. The solid line shows the nominal 95% level.

As pointed in Rukhin & Vangel (1998): *A question of fundamental importance in the analysis of such data is how to form the best estimate of the*

common mean, and what uncertainty to attach to this estimate. The problem is not new in statistical literature, see e.g. Cochran (1937), Graybill & Deal (1959), Fairweather (1972), Paule & Mandel (1982), Jordan & Krishnamoorthy (1996), Yu, Sun & Sinha (1999), Hartung & Makambi (2000), and Witkovský (2001).

For latest developments, in model (1), see Rukhin & Vangel (1998), Rukhin, Biggerstaff & Vangel (2000), Wimmer & Witkovský (2003b), Iyer, Wang & Matthew (2002), and Hartung, Böckenhoff & Knapp (2003).

In this paper, in Section 2, we consider and compare interval estimates for the common mean μ in model (1) based on four methods proposed in recent statistical literature. In particular we will consider interval estimates proposed by Rukhin & Vangel (1998), Hartung, Böckenhoff & Knapp (2003), Iyer, Wang & Matthew (2002), and by Kenward & Roger (1997).

For construction of the interval estimates for the between variance component we propose interval estimator based on the generalized p-values, a method suggested by Tsui & Weerahandi (1989) and Weerahandi (1995). For more details see Wimmer & Witkovský (2003a).

2. Interval estimate of the common mean μ in the model with general between-group variance σ_B^2

If the between laboratory variance σ_B^2 can not be neglected we will consider model (1).

2.1. Rukhin-Vangel method for interval estimation of the common mean μ

Rukhin & Vangel (1998) showed that the computationally simple Mandel-Paule method for estimation of the between-group variance σ_B^2, see Paule & Mandel (1982), can be interpreted as simplified version of the maximum likelihood (ML) method. Further, Rukhin, Biggerstaff & Vangel (2000) showed that the Mandel-Paule estimator of σ_B^2 is usually close to the restricted maximum likelihood (REML) estimator.

The Mandel-Paule estimator of the common mean μ (MP estimator) has the form

$$\hat{\hat{\mu}}^{(MP)} = \frac{\sum_{i=1}^{k} \hat{w}_i^{(MP)} \bar{Y}_i}{\sum_{i=1}^{k} \hat{w}_i^{(MP)}}, \tag{8}$$

where $\hat{w}_i^{(MP)} = 1/(\hat{\sigma}_B^{2(MP)} + S_i^2/n_i)$ are the MP estimators of the weights w_i. Here $\hat{\sigma}_B^{2(MP)}$ estimates the between group variance σ_B^2 and could be derived iteratively from the equation

$$\sum_{i=1}^{k} \frac{\left(\bar{Y}_i - \hat{\hat{\mu}}^{(MP)}\right)^2}{\hat{\sigma}_B^{2(MP)} + S_i^2/n_i} = k - 1. \tag{9}$$

Notice that the left side of (9) is a monotonically decreasing convex function (with probability one), for more details see Paule & Mandel (1982), Rukhin & Vangel (1998) and Iyer, Wang & Mathew (2002).

Rukhin & Vangel (1998) derived consistent estimator of the asymptotic variance of the consensus mean estimator $\hat{\hat{\mu}}^{(MP)}$ (as $k \to \infty$). So, the approximate $(1-\alpha)\times 100\%$ confidence interval proposed by Rukhin & Vangel is given by

$$\hat{\hat{\mu}} \pm z(1-\alpha/2) \sqrt{\sum_{i=1}^{k} \frac{(\bar{Y}_i - \hat{\hat{\mu}})^2}{(\hat{\sigma}_B^2 + S_i^2/n_i)^2}} \Bigg/ \left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}_B^2 + S_i^2/n_i} \right). \tag{10}$$

Here $\hat{\sigma}_B^2$ is the MP estimator or the modified MP estimator (with k instead of $k-1$ in (9)) of σ_B^2 and $\hat{\hat{\mu}}$ is the MP estimator or the modified MP estimator of μ, for more details see Rukhin & Vangel (1998).

In Figure 2 the empirical coverage probabilities of the nominal 95% interval estimator (10) are presented for different situations. As expected, for large number of laboratories k the coverage probabilities are satisfactory and close to the nominal 95% level. However, for medium and small k the actual coverage probabilities are bellow the nominal level. So, we conclude that the Rukhin-Vangel confidence interval (10) for the common mean μ is useful in situations with more than $k = 20$ laboratories.

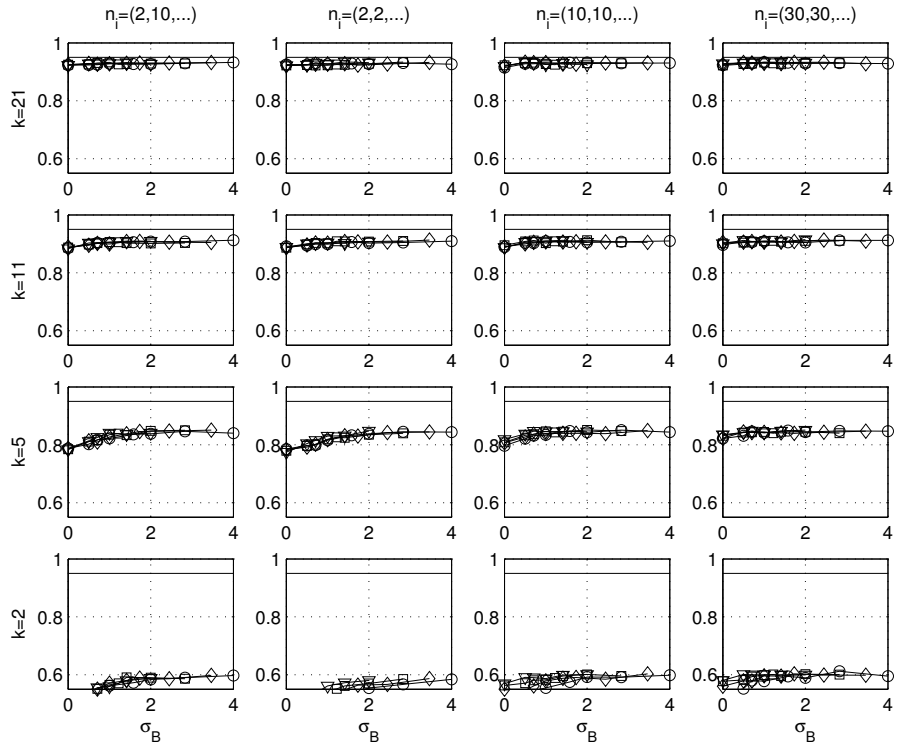

Figure 2. The empirical coverage probabilities of the 95% Rukhin-Vangel confidence intervals (10) calculated with the Mandel-Paule estimates of variance components, based on 10,000 Monte Carlo runs for each specific design as specified in the Appendix A. Here we use the symbol $\bigtriangledown$ for designs with $\sigma_k^2 = 1$, $\square$ for designs with $\sigma_k^2 = 2$, $\diamond$ for designs with $\sigma_k^2 = 3$, and $\bigcirc$ for designs with $\sigma_k^2 = 4$. The solid line shows the nominal 95% level.

2.2. Hartung-Böckenhoff-Knapp method for interval estimation of the common mean μ

Hartung, Böckenhoff & Knapp (2003) proposed another, computationally simple, approximate $(1 - \alpha) \times 100\%$ confidence interval for the common mean μ based on the following considerations.

Let $c_i = w_i/w_\Sigma$, $i = 1, \ldots, k$, where $w_i = 1/(\sigma_B^2 + \sigma_i^2/n_i)$ and $w_\Sigma = \sum_{i=1}^k w_i$. Note that $\hat{\mu}^{(GLS)} = \sum_{i=1}^k c_i \bar{Y}_i$. Further, let

$$U_i^2 = c_i \left(\bar{Y}_i - \sum_{i=1}^{k} c_i \bar{Y}_i \right)^2 \quad \text{and} \quad S^{2(HBK)} = \frac{1}{k-1} \sum_{i=1}^{k} U_i^2. \tag{11}$$

Hartung, Böckenhoff & Knapp (2003) proved that

$$\frac{w_\Sigma}{1 - w_i/w_\Sigma} U_i^2 \sim \chi_1^2, \ i = 1, \ldots, k, \tag{12}$$

$$w_\Sigma \sum_{i=1}^{k} U_i^2 \sim \chi_{k-1}^2, \tag{13}$$

$$\hat{\mu}^{(GLS)} \text{ is independent of } U_i^2, \ i = 1, \ldots, k, \tag{14}$$

$$\frac{\hat{\mu}^{(GLS)} - \mu}{\sqrt{S^{2(HBK)}}} \sim t_{k-1}. \tag{15}$$

Further, they suggested to use estimated weights $\hat{c}_i = \hat{w}_i / \hat{w}_\Sigma$ in (11), where $\hat{w}_i = 1/(\hat{\sigma}_B^2 + \hat{\sigma}_i^2/n_i)$, and $\hat{\sigma}_B^2$ and $\hat{\sigma}_i^2$ are MINQUEs or other quadratic estimators, and showed that

$$\frac{\hat{\hat{\mu}} - \mu}{\sqrt{\hat{S}^{2(HBK)}}} \overset{\text{appr.}}{\sim} t_{k-1}, \tag{16}$$

for more details see Theorem 2.3 and Theorem 4.2 in Hartung, Böckenhoff & Knapp (2003).

So, the approximate $(1 - \alpha) \times 100\%$ confidence interval, proposed by Hartung, Böckenhoff & Knapp is given by

$$\hat{\hat{\mu}} \pm t_{k-1}(1 - \alpha/2) \sqrt{\hat{S}^{2(HBK)}}, \tag{17}$$

where $t_{k-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$-quantile of the Student's t distribution with $k - 1$ degrees of freedom.

In Figure 3 the empirical coverage probabilities of the nominal 95% Hartung-Böckenhoff-Knapp interval estimator (17), calculated with the

Mandel-Paule estimates of variance components, are presented for different situations. The actual coverage probabilities are quite satisfactory for almost all situations where the true between-group variance σ_B^2 is non-negligible. If the true variance σ_B^2 is small, the actual coverage probabilities tends to be bellow the nominal level. In the most extreme situation, considered in our simulations, if $k = 2$ and the number of observations in each laboratory is very small, the actual coverage probabilities are not acceptable and are well bellow the nominal level.

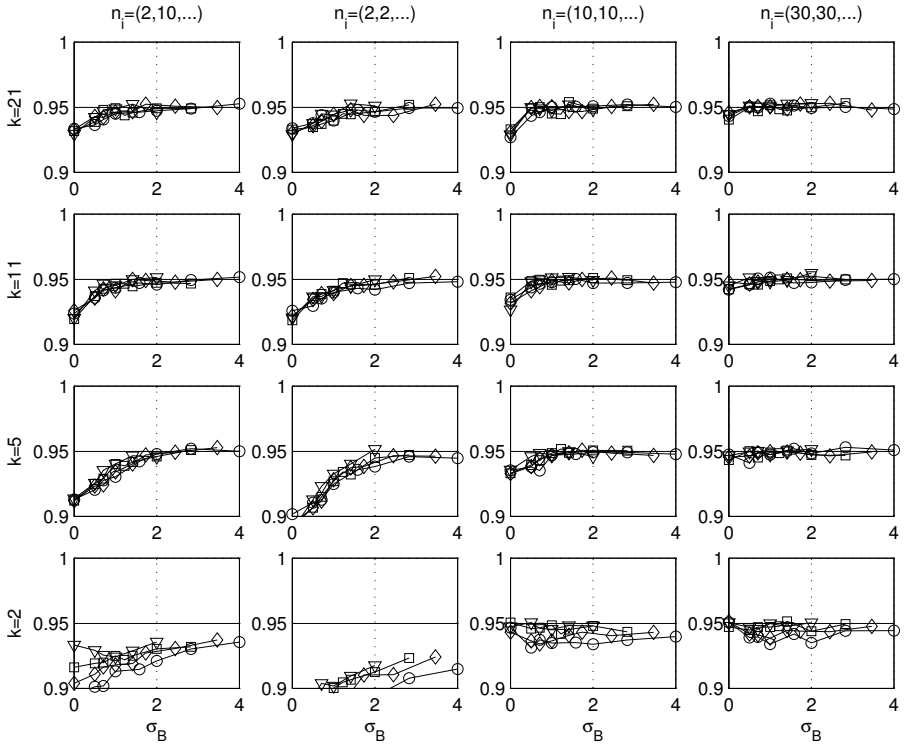

Figure 3. The empirical coverage probabilities of the 95% Hartung-Böckenhoff-Knapp confidence intervals (17) calculated with the Mandel-Paule estimates of variance components, based on 10,000 Monte Carlo runs for each specific design as specified in the Appendix A. Here we use the symbol $\bigtriangledown$ for designs with $\sigma_k^2 = 1$, $\square$ for designs with $\sigma_k^2 = 2$, $\diamond$ for designs with $\sigma_k^2 = 3$, and $\bigcirc$ for designs with $\sigma_k^2 = 4$. The solid line shows the nominal 95% level.

2.3. Iyer-Wang-Mathew method for interval estimation of the common mean μ

Iyer, Wang & Mathew (2002) proposed a confidence interval procedure for the common mean μ using the method of generalized p-values (also known as generalized confidence intervals — GCI), see Tsui and Weerahandi (1989), Weerahandi (1993) and Weerahandi (1995), see also Khuri, Mathew & Sinha (1998).

As the authors showed by large simulation study (the design was similar to that used in this paper), the procedure performs adequately for all sample sizes and the coverage probabilities are very satisfactory. For situations with small between-group variance σ_B^2 the actual coverage probabilities tend to be larger than the nominal 95% level. If the number of laboratories is $k = 2$, the conservativism of the GCI estimator is even more apparent. For more details see Figure 1 in Iyer, Wang & Mathew (2002).

However, the GCI estimate has no explicit expression. To get the limits of the interval estimate of the common mean μ, say $\mu_{Low} \leq \mu \leq \mu_{Upp}$, one should generate a large number, say K (e.g. $K = 10000$), of independent realizations of the proposed generalized pivotal quantity, say R. For more details on the construction of the pivotal quantity R see equation (11) in Iyer, Wang & Mathew (2002).

The pivot R depends functionally on the realized sufficient statistics $\bar{y}_1, \ldots, \bar{y}_k$ and $s_1^2, \ldots, s_k^2$, and on independent random variables $Z \sim N(0,1)$, $Q \sim \chi^2_{k-1}$, and $Q_i \sim \chi^2_{n_i-1}$, $i = 1, \ldots, k$. Its distribution depends only on the parameter of interest, i.e. the common mean μ, and does not depend on the unknown nuisance parameters σ_B^2, and σ_i^2, $i = 1, \ldots, k$. The limits μ_{Low} and μ_{Upp} of the $(1-\alpha) \times 100\%$ GCI estimate are defined by the following equations:

$$\Pr(R \leq \mu_{Low}) = \alpha/2 \quad \text{and} \quad \Pr(R \leq \mu_{Upp}) = 1 - \alpha/2. \tag{18}$$

By generating K independent realizations of R we can estimate the required percentiles of the distribution of R, and hence we can estimate the limits μ_{Low} and μ_{Upp} by the empirical percentiles $R_{[K\alpha/2]}$ and $R_{[K(1-\alpha/2)]}$.

The computation of the proposed GCI is rather intensive as in each step (i.e. for each realization of R) one should calculate the Mandel-Paule solution for the between-group variance.

2.4. Kenward-Roger method for interval estimation of the common mean μ

Based on the results of Harville and his co-workers, see e.g. Kackar & Harville (1984), and Harville & Jeske (1992), Kenward & Roger (1997), suggested a general approach to making small sample inference for fixed effects in general linear mixed effects models by using restricted maximum likelihood (REML) estimates of the unknown variance components.

Savin, Wimmer & Witkovský (2003) proposed to calculate interval estimates for the common mean μ in model (1) by Kenward-Roger method, and as they showed by simulations, the proposed interval estimator has satisfactory coverage probabilities. Moreover, the method is directly available to the researchers by commercial statistical packages, like e.g. SAS.

Here we propose simplified version of the interval estimator based on the Kenward-Roger method. Considering the result of Rukhin, Biggerstaff & Vangel (2000) we suggest to replace the REML estimates of variance components by the Mandel-Paule estimates, otherwise using the explicit formulae for the interval estimator based on the Kenward-Roger method.

In order to derive the explicit formulae we consider now the model (1) in matrix notation:

$$Y = \mu 1_N + Z_0 b + \sum_{i=1}^{k} Z_i \varepsilon_i, \tag{19}$$

where $Y = (Y_{11}, \ldots, Y_{kn_k})'$, $b = (b_1, \ldots, b_k)'$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})'$, for $i = 1, \ldots, k$, and $1_N = (1'_{n_1}, \ldots, 1'_{n_k})'$, $Z_0 = Diag\{1_{n_i}\}$, where $1_{n_i} = (1, \ldots, 1)'$ is n_i-dimensional vector of ones. $Z_i' = [0, \ldots, I_{n_i}, \ldots, 0]$ with I_{n_i} the $(n_i \times n_i)$-dimensional identity matrix, $i = 1, \ldots, k$.

Notice that $b \sim N(0, \sigma_B^2 I_k)$ and $\varepsilon_i \sim N(0, \sigma_i^2 I_{n_i})$ are mutually independent. Under given assumptions

$$Var(Y) = \Sigma = Diag\left\{\sigma_B^2 J_{n_i} + \sigma_i^2 I_{n_i}\right\}, \tag{20}$$

where J_{n_i} denotes the $(n_i \times n_i)$-dimensional matrix of ones.

If the variance components $\vartheta = (\sigma_B^2, \sigma_1^2, \ldots, \sigma_k^2)'$ are known, the GLS estimator of the common mean is given by

$$\hat{\mu}^{(GLS)} = \left(1'_N \Sigma^{-1} 1_N\right)^{-1} 1'_N \Sigma^{-1} Y = \Phi \sum_{i=1}^{k} \frac{n_i \bar{Y}_i}{\sigma_i^2 + n_i \sigma_B^2}, \tag{21}$$

with Φ being its variance

$$\Phi = Var\left(\hat{\mu}^{(GLS)}\right) = \left(1'_N \Sigma^{-1} 1_N\right)^{-1}. \tag{22}$$

If the variance components ϑ are estimated by REML $\hat{\vartheta}$, the plug-in estimator $\hat{\hat{\mu}}^{(REML)}$ of the common mean μ is given by

$$\hat{\hat{\mu}}^{(REML)} = \hat{\Phi} \sum_{i=1}^{k} \frac{n_i \bar{Y}_i}{\hat{\sigma}_i^2 + n_i \hat{\sigma}_B^2}, \tag{23}$$

where $\hat{\Phi} = \Phi(\hat{\vartheta})$.

According to the results of Kenward & Roger (1997) the adjusted estimator of the small sample variance of $\hat{\hat{\mu}}^{(REML)}$ is given by

$$\widehat{Var}\left(\hat{\hat{\mu}}^{(REML)}\right) = \hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda}, \tag{24}$$

where

$$\hat{\Lambda} = \Lambda\left(\hat{\vartheta}\right) = \hat{\Phi}^2 \sum_{i=0}^{k} \sum_{i=0}^{k} \hat{W}_{ij} \left(\hat{Q}_{ij} - \hat{P}_i \hat{\Phi} \hat{P}_j\right), \tag{25}$$

where $\hat{W}_{ij}$, $\hat{Q}_{ij}$, and $\hat{P}_i$, are estimated versions of W_{ij}, Q_{ij}, and P_i defined by

$$\begin{aligned} P_i &= 1'_N \tfrac{\partial \Sigma^{-1}}{\partial \sigma_i^2} 1_N \\ &\text{and} \\ Q_{ij} &= 1'_N \frac{\partial \Sigma^{-1}}{\partial \sigma_i^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j^2} 1_N, \quad i,j = 0,1,\ldots,k, \end{aligned} \tag{26}$$

here we use $\sigma_0^2 = \sigma_B^2$, and W is the asymptotic covariance matrix of $\hat{\vartheta}$, i.e. the inverse of the Fisher information matrix. The explicit formulae are given in the Appendix B.

Here we suggest to use $\hat{\hat{\mu}}^{(MP)}$ instead of $\hat{\hat{\mu}}^{(REML)}$, and MP estimates of variance components instead of the REML $\hat{\vartheta}$. Hence, the approximate $(1-\alpha)\times 100\%$ Kenward-Roger confidence interval for the common mean μ is then given by

$$\hat{\hat{\mu}} \pm \sqrt{\hat{\Phi}_A}\, t_{\hat{m}}(1-\alpha/2), \tag{27}$$

where by $t_{\hat{m}}(1-\alpha/2)$ we denote the $(1-\alpha/2)$-quantile of the Student's t-distribution with $\hat{m}$ degrees of freedom, where $\hat{m}$ is the approximation of the Satterthwaite's degrees of freedom estimator

$$\hat{m} = \frac{2\hat{\Phi}^2}{\widehat{Var}(\hat{\Phi})}, \tag{28}$$

where

$$Var(\hat{\Phi}) = \left(\frac{\partial \Phi}{\partial \vartheta}\right)' W \left(\frac{\partial \Phi}{\partial \vartheta}\right). \tag{29}$$

From that we have

$$\hat{m} = \frac{2}{\hat{\Phi}^2(\hat{P}'\hat{W}\hat{P})}, \tag{30}$$

where $\hat{P} = (\hat{P}_0, \hat{P}_1, \ldots, \hat{P}_k)'$.

In Figure 4 we present the empirical coverage probabilities of the 95% confidence intervals (27), calculated with the Mandel-Paule estimates of the variance components, based on 10,000 Monte Carlo runs for each specific design. The presented simulation study was performed by using our own MATLAB algorithms developed for computing the Kenward-Roger confidence intervals for the common mean. In our calculations, we have used the inverse of the Fisher information matrix as an estimate of the covariance matrix of the variance components estimators, as suggested by Kenward and Roger (1997).

The actual coverage probabilities of the Kenward-Roger confidence interval estimator are very satisfactory for almost all considered situations. If the between-group variance σ_B^2 is small the actual coverage probabilities are slightly above the nominal 95% level (similar pattern observed for the GCI estimator). However, if the number of laboratories is $k = 2$ the actual coverage probabilities are not satisfactory. If the number of observations per laboratory is large, the actual coverage probabilities tends to be well bellow the nominal level.

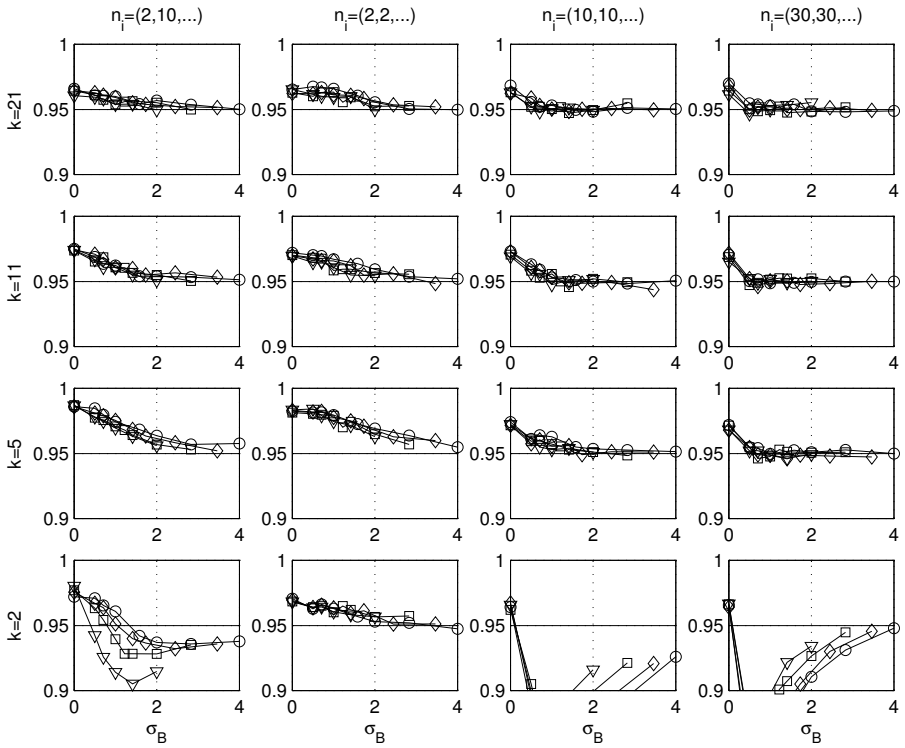

Figure 4. The empirical coverage probabilities of the 95% Kenward-Roger confidence intervals (27) calculated with the Mandel-Paule estimates of variance components, based on 10,000 Monte Carlo runs for each specific design as specified in the Appendix A. Here we use the symbol $\bigtriangledown$ for designs with $\sigma_k^2 = 1$, $\square$ for designs with $\sigma_k^2 = 2$, $\diamond$ for designs with $\sigma_k^2 = 3$, and $\bigcirc$ for designs with $\sigma_k^2 = 4$. The solid line shows the nominal 95% level.

3. Examples

In Table 3 we present the approximate 95% interval estimates for the common mean μ of the Selenium in Non-fat Milk Powder Data, see Table 1, and for the common mean μ of the Arsenic in Oyster Tissue Data, see Table 2, calculated by the four methods e mentioned above. In particular, by the Rukhin-Vangel method with Mandel-Paule estimates

of variance components (RV-MP), by the Hartung-Böckenhoff-Knapp method with Mandel-Paule estimates of variance components (HBK-MP), by the GCI based on the Iyer-Wang-Mathew method (IWM-GCI) with $K = 10000$ independent realizations of the generalized pivotal quantity, and by the Kenward-Roger method with Mandel-Paule estimates of variance components (KR-MP).

Table 3. 95% Interval estimates for common mean calculated by different methods. The variance components were estimated by the Mandel-Paule method.

Data	Method	Equation	$\hat{\hat{\mu}}$	Lower	Upper
Selenium	RV-MP	(10)	109.8214	108.0596	111.5832
Selenium	HBK-MP	(17)	109.8214	105.6741	113.9687
Selenium	IWM-GCI	(18)	109.6798	104.4344	114.6919
Selenium	KR-MP	(27)	109.8214	104.0357	115.6071
Arsenic	RV-MP	(10)	13.2252	12.7095	13.7408
Arsenic	HBK-MP	(17)	13.2252	12.6770	13.7733
Arsenic	IWM-GCI	(18)	13.2265	12.6736	13.7769
Arsenic	KR-MP	(27)	13.2252	12.6749	13.7754

The Selenium data have small number of groups (methods), $k = 4$, they are unbalanced, with medium number of observations per one laboratory. The data show clear heterogeneity of variances. The Mandel-Paule estimate of the between-group variance is $\hat{\sigma}_B^{2(MP)} = 4.1340$, however, the REML estimate is almost zero. According to our simulations we would expect that the correct 95% interval estimate should be a compromise between the HBK-MP (with $\hat{S}^{2(HBK)} = 1.6983$) and KR-MP method (with $\hat{\Phi}_A = 2.1525$ and $\hat{m} = 2.2$), and it is consonant with IWR-GCI method.

The Arsenic data have large number of laboratories, $k = 28$, they are almost balanced, with small number observations per one laboratory. The data show clear heterogeneity of variances. The Mandel-Paule estimate of the between-group variance is $\hat{\sigma}_B^{2(MP)} = 1.9055$, and REML estimate is $\sigma_B^{2(REML)} = 1.9142$. According to our simulations we would expect similar result by all the methods. Here we have $\hat{S}^{2(HBK)} = 0.0714$ for the HBK-MP

method and $\hat{\Phi}_A = 0.0719$ and $\hat{m} = 26.8$ for the KR-MP method. Iyer, Wang & Mathew (2002) reported the 95% generalized confidence interval estimate for Arsenic data as $(12.683, 13.772)$.

4. Conclusions

In this paper we have considered and compared several interval estimators for the common mean μ in general unbalanced heteroscedastic one-way random effects model (1).

We have proposed to use computationally simplified version of the Kenward-Roger confidence interval by using the Mandel-Paule estimators of the variance components. Because of the computational simplicity and based on our simulations we suggest to use this estimator for computing interval estimates of the common mean μ.

As shown by the simulation study, three of the considered interval estimators have satisfactory empirical coverage probabilities in most of the considered situations, namely the Hartung-Böckenhoff-Knapp estimator (HBK-MP), Iyer-Wang-Mathew estimator (IWM-GCI), and the Kenward-Roger estimator (KR-MP).

If the true between-group variance σ_B^2 is small, the actual coverage probabilities of the HBK-MP estimator tend to be bellow the nominal level and those of the IWM-GCI estimator and the KR-MP estimator tend to be above the nominal level.

If the number of laboratories is $k = 2$, the behavior of the estimators is critical. The IWM-GCI estimator tends to be too conservative, i.e. the actual coverage probabilities are above the nominal level, especially if σ_B^2 is small.

The KR-MP estimator has satisfactory coverage probabilities only if the number of observations per laboratory is very small, $n_i = 2$, otherwise the actual coverage probabilities can be well bellow the nominal level.

On the other hand, for $k = 2$, the HBK-MP estimator has satisfactory coverage probabilities for medium and large number of observations per laboratory, but for small number of observations the actual coverage probability can be well bellow the nominal level.

Acknowledgement

The research was supported by the grant from Scientific Grant Agency of the Slovak Republic VEGA 1/0264/03.

Appendix A

The design of the simulation study is similar to that presented in Iyer, Wang & Mathew (2002).

Assuming that model (1) is true, we have used the following values of the unknown parameters in the simulation study: $\mu = 0$, $k \in \{21, 11, 5, 2\}$. Given k, four patterns of n_i were used: $n_i = \{2, 10, 2, 10, \ldots\}$, $i = 1, \ldots, k$, further $n_i = 2$ for all $i = 1, \ldots, k$, $n_i = 10$ for all i and $n_i = 30$ for all i. The within laboratory variances were equally spaced values of $\{\sigma_1^2, \ldots, \sigma_k^2\}$, where $\sigma_1^2 = 1$ and $\sigma_k^2 \in \{1, 2, 3, 4\}$. The values for σ_B^2 were taken to be $\{0, 1/4, 1/2, 1, (1 + \sigma_k^2)/2, \sigma_k^2, 2\sigma_k^2, 4\sigma_k^2\}$.

For each combination of parameters, 10,000 independent realizations of $\bar{y}_1, \ldots, \bar{y}_k$ and $s_1^2, \ldots, s_k^2$ were generated and the 95% confidence interval for μ was calculated.

The relative frequency of cases such that the interval estimate contained the true value $\mu = 0$ was recorded and plotted in the figures.

Appendix B

Notice that

$$\Sigma^{-1} = Diag\left\{\frac{1}{\sigma_i^2}\left(I_{n_i} - \frac{\sigma_B^2}{\sigma_i^2 + n_i\sigma_B^2} J_{n_i}\right)\right\}, \tag{31}$$

$$\det(\Sigma) = \prod_{i=1}^{k} \sigma^{2(n_i-1)}(\sigma_i^2 + n_i\sigma_B^2), \tag{32}$$

and the restricted log-likelihood function for variance components $\vartheta = (\sigma_B^2, \sigma_1^2, \ldots, \sigma_k^2)$ equals to

$$\begin{aligned} loglik(\vartheta) &= -\frac{1}{2}(N-1)\log(2\pi) - \frac{1}{2}\log(\det(\Sigma)) - \frac{1}{2}\log(\det(1_N'\Sigma^{-1}1_N)) \\ &\quad -\frac{1}{2}y'\left(\Sigma^{-1} - \Sigma^{-1}1_N(1_N'\Sigma^{-1}1_N)^{-1}1_N'\Sigma^{-1}\right)y \\ &= -\frac{1}{2}(N-1)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\Big((n_i-1)\log(\sigma_i^2) + \log(\sigma_i^2 + n_i\sigma_B^2)\Big) \end{aligned} \tag{33}$$

$$-\frac{1}{2}\log\left(\sum_{i=1}^{k}\frac{n_i}{\sigma_i^2}(1-\gamma_i)\right)-\frac{1}{2}\sum_{i=1}^{k}\frac{n_i}{\sigma_i^2}\left((n_i-1)\frac{s_i^2}{n_i}+(1-\gamma_i)\bar{y}_i^2\right)$$

$$+\frac{1}{2}\left(\sum_{i=1}^{k}\frac{n_i}{\sigma_i^2}(1-\gamma_i)\bar{y}_i\right)^2\Bigg/\left(\sum_{i=1}^{k}\frac{n_i}{\sigma_i^2}(1-\gamma_i)\right),$$

where $\gamma_i=\sigma_B^2/(\sigma_B^2+\sigma_i^2/n_i)$.

The REML estimator of ϑ is defined as

$$\hat{\vartheta}=\mathrm{argmax}_{\vartheta\in\Theta}loglik(\vartheta), \tag{34}$$

where Θ is the natural parameter space of ϑ.

Lemma 1. *Consider model* (1) *in its matrix form* (19). *Then the following holds true:*

$$\Phi=(1_N'\Sigma^{-1}1_N)^{-1}=\left(\sum_{i=1}^{k}\frac{n_i}{\sigma_i^2+n_i\sigma_B^2}\right)^{-1}, \tag{35}$$

$$P_0=1_N'\frac{\partial\Sigma^{-1}}{\partial\sigma_B^2}1_N=-\sum_{i=1}^{k}\left(\frac{n_i}{\sigma_i^2+n_i\sigma_B^2}\right)^2, \tag{36}$$

$$P_i=1_N'\frac{\partial\Sigma^{-1}}{\partial\sigma_i^2}1_N=-\frac{n_i}{(\sigma_i^2+n_i\sigma_B^2)^2},\quad i=1,\ldots,k, \tag{37}$$

$$Q_{0,0}=1_N'\frac{\partial\Sigma^{-1}}{\partial\sigma_B^2}\Sigma\frac{\partial\Sigma^{-1}}{\partial\sigma_B^2}1_N=\sum_{i=1}^{k}\left(\frac{n_i}{\sigma_i^2+n_i\sigma_B^2}\right)^3, \tag{38}$$

$$Q_{0,i}=1_N'\frac{\partial\Sigma^{-1}}{\partial\sigma_B^2}\Sigma\frac{\partial\Sigma^{-1}}{\partial\sigma_i^2}1_N=Q_{i,0}=\frac{n_i^2}{(\sigma_i^2+n_i\sigma_B^2)^3},\quad i=1,\ldots,k, \tag{39}$$

$$
\begin{aligned}
Q_{ij} &= 1'_N \frac{\partial \Sigma^{-1}}{\partial \sigma_i^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j^2} 1_N \\
&= \begin{cases} 0, & i \neq j, \\ \dfrac{n_i}{(\sigma_i^2 + n_i \sigma_B^2)^3}, & i = j, \end{cases}, \quad i = 1, \ldots, k.
\end{aligned} \tag{40}
$$

The Fisher information matrix for the REML of variance components is defined by its elements as

$$
\begin{aligned}
\{I_F\}_{i,j} &= \frac{1}{2} \left[tr \left(\frac{\partial \Sigma^{-1}}{\partial \sigma_i^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j^2} \Sigma \right) - tr(2 \Phi Q_{ij} - \Phi P_i \Phi P_j) \right] \\
&= \frac{1}{2} \left[\{S\}_{ij} - \{R\}_{ij} \right], \quad i, j = 0, 1, \ldots, k,
\end{aligned} \tag{41}
$$

here we use $\sigma_0^2 = \sigma_B^2$ *and the elements of matrix* R *are given by*

$$
\{R\}_{ij} = \Phi(2Q_{ij} - P_i \Phi P_j), \quad i, j = 0, 1, \ldots, k. \tag{42}
$$

The elements of matrix S *are given by*

$$
\{S\}_{0,0} = -P_0 = \sum_{i=1}^{k} \left(\frac{n_i}{\sigma_i^2 + n_i \sigma_B^2} \right)^2, \tag{43}
$$

$$
\{S\}_{0,i} = \{S\}_{i,0} = -P_i = \frac{n_i}{(\sigma_i^2 + n_i \sigma_B^2)^2}, \quad i = 1, \ldots, k, \tag{44}
$$

$$
\{S\}_{ij} = \begin{cases} 0, & i \neq j, \\ \dfrac{1}{\sigma_i^4} \left(n_i - \dfrac{2 \sigma_B^2 n_i}{\sigma_i^2 + n_i \sigma_B^2} + \dfrac{n_i \sigma_B^4}{(\sigma_i^2 + n_i \sigma_B^2)^2} \right), & i = j, \end{cases}, \; i = 1, \ldots, k. \tag{45}
$$

Proof. The proof is based on the fact that

$$\frac{\partial \Sigma^{-1}}{\partial \sigma_i^2} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i^2} \Sigma^{-1}, \tag{46}$$

and on using (31). ■

Then the asymptotic covariance matrix of $\hat{\vartheta}$, say W, is the inverse of the Fisher information matrix $I_F(\vartheta)$, for more details see Searle, Casella & McCulloch (1992), i.e.

$$W = W(\vartheta) = I_F^{-1}(\vartheta). \tag{47}$$

References

[1] W.G. Cochran, *Problems arising in the analysis of a series of similar experiments*, Journal of the Royal Statistical Society, Supplement **4** (1937), 102–118.

[2] K.R. Eberhardt, C.P. Reeve and C.H. Spiegelman, *A minimax approach to combining means, with practical examples,* Chemometrics Intell. Lab. Systems **5** (1989), 129–148.

[3] W.R. Fairweather, *A method of obtaining an exact confidence interval for the common mean of several normal populations*, Applied Statistics **21** (1972), 229–233.

[4] F.A. Graybill and R.D. Deal, *Combining unbiased estimators*, Biometrics **3** (1959), 1–21.

[5] J. Hartung and K.H. Makambi, *Alternative test procedures and confidence intervals on the common mean in the fixed effects model for meta-analysis*, Technical Report of the Department of Statistics, University of Dortmund 2000.

[6] J. Hartung, A. Böckenhoff and G. Knapp, *Generalized Cochran-Wald statistics in combining of experiments*, Journal of Statistical Planning and Inference **113** (2003), 215–237.

[7] D.A. Harville and D.R. Jeske, *Mean square error of estimation or prediction under a general linear model*, Journal of the American Statistical Association **87** (1992), 724–731.

[8] H.K. Iyer, C.M. Wang and T. Mathew, *Models and confidence intervals for true values in interlaboratory trials*, Manuscript submitted for publication 2002.

[9] S.M. Jordan and K. Krishnamoorthy, *Exact confidence intervals for the common mean of several normal populations*, Biometrics **52** (1996), 77–86.

[10] A.N. Kackar and D.A. Harville, *Approximations for standard errors of estimators of fixed and random effects in mixed linear models*, Journal of the American Statistical Association **79** (1984), 853–862.

[11] M.G. Kenward and J.H. Roger, *Small sample inference for fixed effects from restricted maximum likelihood*, Biometrics **53** (1997), 983–997.

[12] A.I. Khuri, T. Mathew and B.K. Sinha, *Statistical Tests for Mixed Linear Models*, J. Wiley & Sons, New York 1998.

[13] R.C. Paule and J. Mandel, *Consensus values and weighting factors*, Journal of Research of the National Bureau of Standards **87** (5) (1982), 377–385.

[14] A.L. Rukhin, B.J. Biggerstaff and M.G. Vangel, *Restricted maximum likelihood estimation of a common mean and the Mandel-Paule algorithm*, Journal of Statistical Planning and Inference **83** (2000), 319–330.

[15] A.L. Rukhin and M.G. Vangel, *Estimation of a common mean and weighted means statistics*, Journal of the American Statistical Association **93** (441) (1998), 303–308.

[16] A. Savin, G. Wimmer and V. Witkovský, *On Kenward-Roger confidence intervals for common mean in interlaboratory trials*, Measurement Science Review **3** Section 1, (2003), 53–56. http://www.measurement.sk/.

[17] S.R. Searle, G. Casella and C.E. McCulloch, Variance Components, J. Wiley & Sons, New York 1992.

[18] K.W. Tsui and S. Weerahandi, *Generalized p values in significance testing of hypotheses in the presence of nuisance parameters*, Journal of the American Statistical Association **84** (1989), 602–607.

[19] S. Weerahandi, *Generalized confidence intervals*, Journal of the American Statistical Association **88** (1993), 899–905.

[20] S. Weerahandi, *Exact Statistical Methods for Data Analysis*, Springer-Verlag, New York 1995.

[21] G. Wimmer and V. Witkovský, *Between group variance component interval estimation for the unbalanced heteroscedastic one-way random effects model*, Journal of Statistical Computation and Simulation **73** (2003a), 333–346.

[22] G. Wimmer and V. Witkovský, *Consensus mean and interval estimators for the common mean*, ProbaStat 2002, Submitted to Tatra Mountains Mathematical Publications, (2003b).

[23] V. Witkovský, *On the exact computation of the density and of the quantiles of linear combinations of t and F random variables*, Journal of Statistical Planning and Inference **94** (2001), 1–13.

[24] L.H. Yu, Y. Sun and B.K. Sinha, *On exact confidence intervals for the common mean of several normal populations*, Journal of Statistical Planning and Inference **81** (1999), 263–277.

Received 9 January 2003
Revised 9 December 2003