Tatra Mt. Math. Publ. **26** (2003), 183–194

TATRA MOUNTAINS
Mathematical Publications

CONSENSUS MEAN AND INTERVAL ESTIMATORS FOR THE COMMON MEAN

VIKTOR WITKOVSKÝ — GEJZA WIMMER

ABSTRACT. We consider the problem of interlaboratory studies where the number of measurements made at each laboratory may differ, and, in general, the laboratories may exhibit the between laboratory variability caused by the systematic error due to each laboratory, as well as different within laboratory variances caused by different within laboratory precisions of the measurement methods.

In the case when the between laboratory variability can be neglected, we propose two generalized confidence interval estimators (GPV_1 and GPV_2) for the common mean based on the generalized p-value approach introduced by Tsui and Weerahandi [*Generalized p values in significance testing of hypotheses in the presence of nuisance parameters*, J. Amer. Statist. Assoc. **84** (1989), 602–330] and Weerahandi [*Exact Statistical Methods for Data Analysis*, Springer-Verlag, New York, 1995]. The applicability of the suggested estimators is illustrated by a numerical example where the proposed estimators are compared with other two methods: the FW method, see Fairweather [*A method of obtaining an exact confidence interval for the common mean of several normal populations*, Applied Statistics **21** (1972), 229–233]. and the JK method, see Jordan and Krishnamoorthy [*Exact confidence intervals for the common mean of several normal populations*, Biometrics **52**, 1996, 77–86]. The coverage probabilities of the proposed interval estimators were examined using simulations. During the simulation study we have observed that the GPV_1 interval estimator tends to be anti-conservative and the GPV_2 generalized confidence interval estimator tends to be conservative. Thus GPV_1 results in shorter intervals while GPV_2 leads to longer intervals.

1. Introduction

Assume that the measurements of virtually the same object of interest are made by $k \geq 2$ laboratories. The ith laboratory repeats its measurements n_i times, $n_i \geq 2$. In general, the laboratories may exhibit the between laboratory

2000 Mathematics Subject Classification: Primary 62F25; Secondary 62N99.
Keywords: interlaboratory study, generalilzed p-values, random effects model, heteroscedasticity, consensus mean.

Research for this paper was supported by Grant VEGA 1/7295/20 from the Science Grant Agency of the Slovak Academy of Sciences.

variability, as well as different within laboratory variances (heteroscedasticity). It is assumed that the measurements follow normal distribution. In this paper we will concentrate on the case when the between laboratory variability can be neglected.

The results of a typical interlaboratory study (given in aggregated form) are presented in Table 1: In Eberhardt *et al.* (1989) the data on selenium in non-fat milk powder were reported. The measurements are based on four independent measurement methods.

TABLE 1. Selenium in non-fat milk powder data. Source: Eberhardt *et al.* (1989).

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

As pointed out in Rukhin and Vangel (1998): *A question of fundamental importance in the analysis of such data is how to form the best consensus mean, and what uncertainty to attach to this estimate.* This fundamental question is followed by a series of other questions regarding the statistical inference on the unknown common mean. The problem, although not new in statistical literature, see, e.g., Cochran (1937), Graybill and Deal (1959), is not completely solved. For recent developments in this field see Paule and Mandel (1982), Rukhin and Vangel (1998), Yu *et al.* (1999), Rukhin *et al.* (2000), Rukhin (2002), Iyer *et al.* (2002), Wimmer and Witkovský (2003).

2. Exact and approximate confidence intervals for the common mean

A basic tool for analysis of interlaboratory measurements is typically the one-way random effects model, which may be both unbalanced and heteroscedastic, that is

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

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

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

However, here we will assume that the between laboratory variability can be neglected, that is $\sigma_B^2 = 0$, i.e.,

$$Y_{ij} = \mu + \varepsilon_{ij}, \tag{2}$$

$\varepsilon_{ij} \sim N(0, \sigma_i^2)$ for $i = 1, \dots, k$ and $j = 1, \dots, n_i$. One way to check that $\sigma_B^2 = 0$ is to test the null hypothesis $H_0\colon \sigma_B^2 = 0$ against the one-sided alternative $H_1\colon \sigma_B^2 > 0$ by a generalized significance test in the model (1), or to construct the generalized confidence interval estimate for σ_B^2. For more details see Wimmer and Witkovský (2003).

Yu *et al.* (1999) discussed and compared several *exact* confidence intervals for the common mean based on pivots and p-values in the model (2).

Let $\bar{Y}_i = (1/n_i) \sum_{j=1}^{n_i} Y_{ij}$, and $S_i^2 = \big(1/(n_i - 1)\big) \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$, be the sample laboratory means and the sample laboratory variances. Observe that

$$\bar{Y}_i \sim N(\mu, \sigma_i^2/n_i), \quad \text{and} \quad (n_i - 1)S_i^2 \sim \sigma_i^2 \chi^2_{n_i - 1} \tag{3}$$

are mutually independent random variables.

From that, given the true value of the common mean μ, we get independent t or F random variables

$$t_i(\mu) = \frac{\sqrt{n_i}(\bar{Y}_i - \mu)}{S_i} \sim t_{n_i - 1}, \qquad i = 1, \dots, k, \tag{4}$$

or

$$F_i(\mu) = \frac{n_i(\bar{Y}_i - \mu)^2}{S_i^2} \sim F_{1, n_i - 1}, \qquad i = 1, \dots, k. \tag{5}$$

Fairweather (1972) suggested the exact confidence interval for μ using a pivot based on the weighted linear combination of the independent t variables, i.e.,

$$W_t(\mu) = \sum_{i=1}^{k} u_i t_i(\mu), \tag{6}$$

for some choice of weights $u_i \geq 0$, $\sum_{i=1}^{k} u_i = 1$. Fairweather suggested to use the weights u_i inversely proportional to the variances of random variables $t_i(\mu)$.

If $b(\alpha/2)$ denotes the upper cut-off point of the distribution of $W_t(\mu)$,

$$\Pr\big(|W_t(\mu)| \leq b(\alpha/2) | \mu\big) = 1 - \alpha, \tag{7}$$

then the exact (symmetric, two-sided) $100(1-\alpha)$ % confidence interval for the common mean μ is

$$\left[\frac{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i\bar{Y}_i - b(\alpha/2)}{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i},\ \frac{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i\bar{Y}_i + b(\alpha/2)}{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i}\right] \tag{8}$$

with the consensus mean (the center of interval estimator) given by

$$\hat{\mu}^{(FW)} = \frac{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i\bar{Y}_i}{\sum_{i=1}^{k}\big(\sqrt{n_i}/S_i\big)u_i}\,. \tag{9}$$

Note that the consensus mean and interval estimator critically depend on the choice of the weights u_i.

Jordan and Krishnamoorthy (1996) suggested the exact confidence interval for μ using a pivot based on the weighted linear combination of the independent F variables, i.e.,

$$W_F(\mu) = \sum_{i=1}^{k} v_i F_i(\mu)\,. \tag{10}$$

If $c(\alpha)$ denotes the upper cut-off point of the distribution of $W_F(\mu)$,

$$\Pr\big(W_F(\mu) \le c(\alpha)\,|\,\mu\big) = 1-\alpha\,, \tag{11}$$

then the exact (symmetric, two-sided) $100(1-\alpha)$ % confidence interval for the common mean μ is

$$\left[\sum_{i=1}^{k} p_i\bar{Y}_i - \Delta\,,\ \sum_{i=1}^{k} p_i\bar{Y}_i + \Delta\right] \tag{12}$$

with the consensus mean (the center of interval estimator) given by

$$\hat{\mu}^{(JK)} = \sum_{i=1}^{k} p_i\bar{Y}_i\,, \tag{13}$$

where

$$p_i = \frac{v_i(n_i/S_i^2)}{\sum_{j=1}^{k} v_j(n_j/S_j^2)} \tag{14}$$

and

$$\Delta^2 = \frac{c(\alpha)}{\sum_{j=1}^{k} v_j (n_j/S_j^2)} - \left\{ \sum_{i=1}^{k} p_i \bar{Y}_i^2 - \left(\sum_{i=1}^{k} p_i \bar{Y}_i \right)^2 \right\}. \tag{15}$$

As before, the consensus mean and interval estimator critically depend on the choice of the weights v_i.

For the methods of computing the distribution and the critical values of linear combinations of independent t and F variables see Witkovský (2001).

Exact confidence intervals based on the p-values suggested by Yu *et al.* (1999) are similar in nature. Let $F_{i,\mathrm{obs}}(\mu) = n_i(\bar{y}_i - \mu)^2/s_i^2$ be the observed value of the $F_i(\mu)$ random variable, where $\bar{y}_i$ is the observed value of $\bar{Y}_i$ and s_i^2 is the observed value of S_i^2. Then the ith p-value is given by

$$P_i(\mu) = \Pr\big(F_i(\mu) > F_{i,\mathrm{obs}}(\mu) \,|\, \mu\big). \tag{16}$$

For instance, the exact confidence interval based on Fisher's method is obtained by inverting the acceptance region of the test of the null hypothesis $H_0\colon \mu = \mu_0$.

This test rejects the null hypothesis when

$$-2 \sum_{i=1}^{k} \log P_i(\mu_0) > \chi^2_{2k}(\alpha), \tag{17}$$

where $\chi^2_{2k}(\alpha)$ denotes the upper cut-off point of the χ^2-distribution with $2k$ degrees of freedom, and the $100(1-\alpha)$ % confidence interval is given by

$$\left\{ \mu\colon\ -2 \sum_{i=1}^{k} \log P_i(\mu) \le \chi^2_{2k}(\alpha) \right\}. \tag{18}$$

As before, the interval estimator critically depends on the particular choice of the weighting of the p-values $P_i(\mu)$.

3. Consensus mean based on the generalized p-values

Tsui and Weerahandi (1989), see also Weerahandi (1995), suggested a new concept for (significance) testing hypotheses and interval estimation of the parameters of interest in models with nuisance parameters.

Briefly, the principle lies in the construction of the generalized test variables (or the generalized pivots), say $T(X, x, \theta, \vartheta)$, which functionally depend on the random vector X, its observed value x, the parameter of interest θ and the vector of nuisance parameters ϑ—the random variable $T(X, x, \theta, \vartheta)$ should be such that

- $T_{\mathrm{obs}}(\theta) = T(x, x, \theta, \vartheta)$ does not depend on the unknown nuisance parameters,
- if θ is the true value of the parameter, the probability distribution of $T(X, x, \theta, \vartheta)$ is free of nuisance parameters,
- for fixed x and ϑ, the variables $T(X, x, \theta, \vartheta)$ are stochastically ordered.

Here we will define two potential generalized test variables for constructing the consensus mean and interval estimates for the common mean. The first one we define as the *contrast variable*

$$T_1(\mu) = T_1\left([\bar{Y}_1, \dots, \bar{Y}_k; S_1^2, \dots, S_k^2], [\bar{y}_1, \dots, \bar{y}_k; s_1^2, \dots, s_k^2], \mu, [\sigma_1^2, \dots, \sigma_k^2]\right), \tag{19}$$

given as

$$T_1(\mu) = \left(\sum_{i=1}^{k} \frac{(\bar{Y}_i - \mu)^2}{\sigma_i^2/n_i} \right) \Bigg/ \left(\sum_{i=1}^{k} \frac{(\bar{y}_i - \mu)^2}{\left(\frac{\sigma_i^2}{n_i} \frac{s_i^2(n_i - 1)}{S_i^2(n_i - 1)} \right)} \right). \tag{20}$$

Note that $T_{1,\mathrm{obs}}(\mu) = 1$, and if μ is the true value of the common mean in the model (2), then the distribution of $T_1(\mu)$ does not depend on the nuisance parameters σ_i^2, $i = 1, \dots, k$, and is given as

$$T_1(\mu) \sim U \Bigg/ \left(\sum_{i=1}^{k} \frac{(\bar{y}_i - \mu)^2}{\left(\frac{(n_i - 1)}{n_i} \frac{s_i^2}{U_i} \right)} \right), \tag{21}$$

where $U \sim \chi_k^2$ and $U_i \sim \chi_{n_i - 1}^2$ denote the mutually independent chi-square distributed random variables with k and $n_i - 1$, $i = 1, \dots, k$, degrees of freedom, respectively.

The null hypothesis $H_0 \colon \mu = \mu_0$ is rejected by the significance test based on the generalized p-value if

$$\begin{aligned} P_1(\mu_0) &= \Pr\big(T_1(\mu) > T_{1,\mathrm{obs}}(\mu) | \mu = \mu_0\big) \\ &= \Pr\left(U > \sum_{i=1}^{k} \frac{(\bar{y}_i - \mu_0)^2}{(n_i - 1) s_i^2 / n_i} U_i \right) < \alpha, \end{aligned} \tag{22}$$

where α is the chosen significance level. Note that the p-value (22) can be evaluated exactly by (widely available) Imhof's algorithm, see Imhof (1961), or simply by simulating the distribution of the event $T_1(\mu_0) > 1$ given μ_0 and the observed data $\bar{y}_1, \dots, \bar{y}_k$ and $s_1^2, \dots, s_k^2$.

As the second potential generalized test variable (the improved version of (20) which takes into account full information available from the data) we suggest the second *contrast variable*

$$T_2(\mu) = \left(\sum_{i=1}^{k} \frac{(\bar{Y}_i - \mu)^2}{\sigma_i^2/n_i}\right) \Bigg/ \left(\sum_{i=1}^{k} \frac{(\bar{y}_i - \mu)^2}{\left(\dfrac{\sigma_i^2}{n_i}\dfrac{s_i^2(n_i-1) + n_i(\bar{y}_i - \mu)^2}{S_i^2(n_i-1) + n_i(\bar{Y}_i - \mu)^2}\right)}\right). \tag{23}$$

Note that $T_{2,\text{obs}}(\mu) = 1$, and the distribution of $T_2(\mu)$ does not depend on the nuisance parameters σ_i^2, and is given as

$$T_2(\mu) \sim \sum_{i=1}^{k} V_i \Bigg/ \left(\sum_{i=1}^{k} \frac{(\bar{y}_i - \mu)^2}{(n_i-1)s_i^2/n_i + (\bar{y}_i - \mu)^2}(U_i + V_i)\right), \tag{24}$$

where $V_i \sim \chi_1^2$ and $U_i \sim \chi_{n_i-1}^2$, $i = 1, \ldots, k$, denote the mutually independent chi-square distributed random variables with 1 and $n_i - 1$ degrees of freedom, respectively.

The null hypothesis $H_0: \mu = \mu_0$ is rejected by the significance test based on the generalized p-value if

$$\begin{aligned} P_2(\mu_0) &= \Pr(T_2(\mu) > T_{2,\text{obs}}(\mu) | \mu = \mu_0) && (25) \\ &= \Pr\left(\sum_{i=1}^{k} \frac{(n_i-1)s_i^2/n_i}{(n_i-1)s_i^2/n_i + (\bar{y}_i - \mu_0)^2} V_i > \sum_{i=1}^{k} \frac{(\bar{y}_i - \mu_0)^2}{(n_i-1)s_i^2/n_i + (\bar{y}_i - \mu_0)^2} U_i\right) < \alpha, \end{aligned}$$

where α is the chosen significance level. Again, the p-value (25) can be evaluated exactly by Imhof's algorithm.

Now we can define the consensus mean based on the generalized p-values (22) and (25)—the GPV consensus mean, as

$$\hat{\mu}^{(\mathrm{GPV}_i)} = \operatorname{argmax}_{\mu_0} P_i(\mu_0), \qquad i = 1, 2. \tag{26}$$

Given the observed values $\bar{y}_1, \ldots, \bar{y}_k$ and $s_1^2, \ldots, s_k^2$, we define interval estimate for the common mean μ (the generalized confidence interval) as the acceptance region of the significance test based on the generalized p-values (22) and (25) of the hypothesis $H_0: \mu = \mu_0$,

$$CI^{(\mathrm{GPV}_i)} = \{\mu_0 : P_i(\mu_0) \geq \alpha\}, \qquad i = 1, 2, \tag{27}$$

where α is the chosen significance level.

4. Example

For illustration of some of the mentioned methods, we use the selenium in non-fat milk powder data given in Table 1.

First we should check if we can neglect the between laboratory variability. We will do that by testing the null hypothesis $H_0\colon \sigma_B^2 = 0$ against the one-sided alternative $H_1\colon \sigma_B^2 > 0$ by a generalized significance test in the model (1) For more details see W i m m e r and W i t k o v s k ý (2003).

For a chosen significance level α, the significance test will reject H_0 if

$$P_3(0) = \Pr\big(T_3(\sigma_B^2) > T_{3,\mathrm{obs}}(\sigma_B^2) | \sigma_B^2 = 0\big) < \alpha\,, \tag{28}$$

where the generalized test variable $T_3(\sigma_B^2)$ is defined as

$$T_3(\sigma_B^2) = \sum_{i=1}^{k} \frac{1}{\sigma_B^2 + \sigma_i^2/n_i} \left(\bar{Y}_i - \frac{\sum\limits_{j=1}^{k} \bar{Y}_j/(\sigma_B^2 + \sigma_j^2/n_j)}{\sum\limits_{j=1}^{k} 1/(\sigma_B^2 + \sigma_j^2/n_j)} \right)^2 \Bigg/$$

$$\sum_{i=1}^{k} \frac{1}{\sigma_B^2 + \frac{\sigma_i^2}{n_i} \frac{(n_i-1)}{(n_i-1)} \frac{s_i^2}{S_i^2}} \left(\bar{y}_i - \frac{\sum\limits_{j=1}^{k} \bar{y}_j \Big/ \Big(\sigma_B^2 + \frac{\sigma_j^2}{n_j} \frac{(n_j-1)}{(n_j-1)} \frac{s_j^2}{S_j^2}\Big)}{\sum\limits_{j=1}^{k} 1 \Big/ \Big(\sigma_B^2 + -\frac{\sigma_j^2}{n_j} \frac{(n_j-1)}{(n_j-1)} \frac{s_j^2}{S_j^2}\Big)} \right)^2 . \tag{29}$$

Note that $T_{3,\mathrm{obs}}(\sigma_B^2) = 1$, and the distribution of $T_3(\sigma_B^2)$, given the true value of the between laboratory variance σ_B^2, does not depend on the nuisance parameters μ and σ_i^2, and is given as

$$T_3(\sigma_B^2) \sim U \Bigg/ \sum_{i=1}^{k} \frac{1}{\sigma_B^2 + \frac{(n_i-1)}{n_i} \frac{s_i^2}{U_i}} \left(\bar{y}_i - \frac{\sum\limits_{j=1}^{k} \bar{y}_j \Big/ \Big(\sigma_B^2 + \frac{(n_j-1)}{n_j} \frac{s_j^2}{U_j}\Big)}{\sum\limits_{j=1}^{k} 1 \Big/ \Big(\sigma_B^2 + \frac{(n_j-1)}{n_j} \frac{s_j^2}{U_j}\Big)} \right)^2 , \tag{30}$$

where $U \sim \chi_k^2$ and $U_i \sim \chi_{n_i-1}^2$ denote the mutually independent chi-square distributed random variables with k and $n_i - 1$, $i = 1, \dots, k$, degrees of freedom, respectively.

For given data from Table 1, using 1,000,000 Monte-Carlo runs to generate realizations of U and U_i, $i = 1, \dots, k$, we have estimated the generalized p-value (28) to be $P_3(0) = 0.2093$. So, at the significance level $\alpha = 0.05$ we cannot reject the null hypothesis $H_0\colon \sigma_B^2 = 0$. For the sake of simplicity, we adopt the model (2) for point and interval estimation of the common mean of selenium in non-fat milk powder data.

Point and interval estimates of the common mean based on four different methods are presented in Table 2.

TABLE 2. Selenium in non-fat milk powder data. The consensus mean and interval estimate for the common mean μ calculated by four different methods.

Method	Equations	Consensus Mean $\bar{\mu}$	95 % Interval Estimate
FW	(8), (9)	109.6546	$\langle 108.5369;\ 110.7722\rangle$
JK	(12), (13)	109.5590	$\langle 108.4588;\ 110.6621\rangle$
GPV_1	(22), (26), (27)	109.5824	$\langle 108.6178;\ 110.5698\rangle$
GPV_2	(25), (26), (27)	109.5480	$\langle 108.5225;\ 110.6799\rangle$

5. Simulation study

In our simulation study we have examined the coverage probabilities of the proposed generalized confidence interval estimators for the common mean μ in model (2). The design of the simulation study is similar to that presented in Iyer *et al.* (2002). Assuming that model (2) is true, we have used the following values of the unknown parameters in the simulation study: $\mu = 0$, $k \in \{2, 5, 11, 21\}$. Given k, four patterns of n_i were used: $n_i = 2$ for all $i = 1, \dots, k$, $n_i = 10$ for all i, $n_i = 30$ for all i, and $n_i = \{2, 10, 2, 10, \dots\}$. The within laboratory variances were equally spaced values of $\{\sigma_1^2, \dots, \sigma_k^2\}$, where $\sigma_1^2 = 1$ and $\sigma_k^2 \in \{1, 4, 25, 100\}$. For each combination of parameters, 10,000 independent realizations of $\bar{y}_1, \dots, \bar{y}_k$ and $s_1^2, \dots, s_k^2$ were generated and the generalized p-values were calculated using Imhof's algorithm and compared with the chosen significance level $\alpha = 0.05$.

During the simulation study we observed that the GPV_1 generalized confidence interval estimator is anti-conservative, i.e., the coverage probability is typically less than the nominal $1 - \alpha$, and the GPV_2 generalized confidence interval estimator is conservative one, with the coverage probabilities greater than nominal $1 - \alpha$. This results in shorter intervals for GPV_1 and in longer intervals for GPV_2.

Based on that observation, in Table 3 we present just the relative number of cases when $P_2(0) \geq \alpha$, the empirical coverage probability of the proposed GPV_2 95 % interval estimator for the common mean μ, see (25) and (27).

6. Conclusions

We have proposed two alternative point and interval estimators of the common mean—the problem motivated by interlaboratory studies. The proposed estimators are applicable if we can neglect the between laboratory variance and hence, to adopt a simpler model. We have also suggested a method for

TABLE 3. Coverage probability of the 95% generalized confidence interval estimator GPV_2 for the common mean μ.

$n_i = 2$ for all $i = 1, \dots, k$

	$\sigma_k^2 = 1$	$\sigma_k^2 = 4$	$\sigma_k^2 = 25$	$\sigma_k^2 = 100$
$k = 2$	0.9630	0.9628	0.9674	0.9670
$k = 5$	0.9735	0.9718	0.9687	0.9712
$k = 11$	0.9769	0.9774	0.9747	0.9771
$k = 21$	0.9756	0.9736	0.9754	0.9752

$n_i = 10$ for all $i = 1, \dots, k$

	$\sigma_k^2 = 1$	$\sigma_k^2 = 4$	$\sigma_k^2 = 25$	$\sigma_k^2 = 100$
$k = 2$	0.9649	0.9663	0.9695	0.9643
$k = 5$	0.9897	0.9887	0.9886	0.9890
$k = 11$	0.9947	0.9951	0.9949	0.9952
$k = 21$	0.9966	0.9966	0.9956	0.9966

$n_i = 30$ for all $i = 1, \dots, k$

	$\sigma_k^2 = 1$	$\sigma_k^2 = 4$	$\sigma_k^2 = 25$	$\sigma_k^2 = 100$
$k = 2$	0.9527	0.9558	0.9549	0.9570
$k = 5$	0.9611	0.9613	0.9601	0.9579
$k = 11$	0.9590	0.9619	0.9627	0.9593
$k = 21$	0.9618	0.9622	0.9550	0.9605

$n_i = \{2, 10, 2, 10, \dots\}$

	$\sigma_k^2 = 1$	$\sigma_k^2 = 4$	$\sigma_k^2 = 25$	$\sigma_k^2 = 100$
$k = 2$	0.9655	0.9666	0.9630	0.9670
$k = 5$	0.9771	0.9767	0.9788	0.9787
$k = 11$	0.9825	0.9840	0.9829	0.9822
$k = 21$	0.9837	0.9838	0.9834	0.9860

significance testing of the null hypothesis $H_0: \sigma_B^2 = 0$ against one-sided alternative, again, based on the generalized p-values. If the null hypothesis is rejected, other methods of interval estimation should be considered, e.g., the generalized confidence interval proposed by Iyer *et al.* (2002).

The p-values required for construction of the proposed estimators can be evaluated exactly (numerically) by Imhof's algorithm. Based on our simulation study we conclude that the properties of the proposed interval estimators are not optimal. The GPV_1 interval estimator shows anti-conservativism while the GPV_2 interval estimator is conservative from the frequentist point of view. However, the presented real data example illustrates the applicability of the proposed estimators as well as their comparison with two other methods. Comparison with other methods was not considered in this paper and is a matter of further studies.

REFERENCES

[1] COCHRAN, W. G.: *Problems arising in the analysis of a series of similar experiments*, J. Roy. Statist. Soc. Supplement **4** (1937), 102–118.

[2] EBERHARDT, K. R.—REEVE, C. P.—SPIEGELMAN, C. H.: *A minimax approach to combining means, with practical examples*, Chemometrics Intell. Lab. Systems **5** (1989), 129–148.

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

[4] GRAYBILL, F. A.—DEAL, R. D.: *Combining unbiased estimators*, Biometrics **3** (1959), 1–21.

[5] IMHOF, J. P.: *Computing the distribution of quadratic forms in normal variables*, Biometrika **48** (1961), 481–482.

[6] IYER, H. K.—WANG, C. M.—MATHEW, T.: *Models and confidence intervals for true values in interlaboratory trials*, J. Amer. Statist. Assoc., (submitted).

[7] JORDAN, S. M.—KRISHNAMOORTHY, K.: *Exact confidence intervals for the common mean of several normal populations*, Biometrics **52** (1996), 77–86.

[8] PAULE, R. C.—MANDEL, J.: *Consensus values and weighting factors*, J. Res. Nat. Bur. Stand. **87** (1982), 377–385.

[9] RUKHIN, A. L.: *Two procedures of meta-analysis in clinical trials and interlaboratory studies*, Tatra Mt. Math. Publ. **26** (2003), 155–168.

[10] RUKHIN, A. L.—VANGEL, M. G.: *Estimation of a common mean and weighted means statistics*, J. Amer. Statist. Assoc. **93** (1998), 303–308.

[11] RUKHIN, A. L.—BIGGERSTAFF, B. J.—VANGEL, M. G.: *Restricted maximum likelihood estimation of a common mean and the Mandel-Paule algorithm*, J. Statist. Plann. Inference **83** (2000), 319–330.

[12] TSUI, K. W.—WEERAHANDI, S.: *Generalized p values in significance testing of hypotheses in the presence of nuisance parameters*, J. Amer. Statist. Assoc. **84** (1989), 602–607.

[13] WEERAHANDI, S.: *Exact Statistical Methods for Data Analysis*, Springer-Verlag, New York, 1995.

[14] WIMMER, G.—WITKOVSKÝ, V.: *Between group variance component interval estimation for the unbalanced heteroscedastic one-way random effects model*, J. Statist. Comput. Simulation **73** (2003), 333–346.

[15] WITKOVSKÝ, V.: *On the exact computation of the density and of the quantiles of linear combinations of t and F random variables*, J. Statist. Plann. Inference **94** (2001), 1–13.

[16] YU, L. H.—SUN, Y.—SINHA, B. K.: *On exact confidence intervals for the common mean of several normal populations*, J. Statist. Plann. Inference **81** (1999), 263–277.

Received February 4, 2002

Viktor Witkovský
Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9
SK–841 04 Bratislava
SLOVAKIA
E-mail: umerwitk@savba.sk

Gejza Wimmer
Mathematical Institute
Slovak Academy of Sciences
Štefánikova 49
SK–814 73 Bratislava
SLOVAKIA

Matej Bel University
Tajovského 40
SK–974 00 Banská Bystrica
SLOVAKIA
E-mail: wimmer@mat.savba.sk