On Kenward-Roger Confidence Intervals for Common Mean in Interlaboratory Trials[1]

[a1]**Alexander Savin,** [b]**Gejza Wimmer, and** [a2]**Viktor Witkovský**[2]

[a1,2]Institute of Measurement Science, Slovak Academy of Sciences
Dúbravská cesta 9, 841 04 Bratislava, Slovak Republic
[a1]E-mail: savin@savba.sk
[a2]E-mail: witkovsky@savba.sk

[b]Faculty of Natural Sciences,Matej Bel University
Tajovského 40, 974 01 Banská Bystrica, Slovak Republic, and
Mathematical Institute, Slovak Academy of Sciences
Štefánikova 49, 814 73 Bratislava, Slovak Republic
E-mail: wimmer@mat.savba.sk

Abstract. *In this paper we consider the Kenward-Roger method for interval estimation of the common mean — the problem which appears in interlaboratory studies but is also closely related to the multicenter clinical trials and meta-analysis. Here we present the statistical properties (based on large simulation study) of the proposed Kenward-Roger confidence intervals. The method is also illustrated by two examples.*

Keywords: *Interlaboratory study; Common mean; Kenward-Roger Confidence Interval.*

1. Introduction

In interlaboratory trials we consider that the measurements on virtually the same object of interest are made by $k \geq 2$ laboratories. The ith laboratory repeats its measurements n_i times, $n_i \geq 2$. The laboratories may exhibit the between laboratory variability, as well as different within-laboratory variances (heteroscedasticity). Here we will assume that the measurements follow normal distribution.

We consider the following random effects model:

$$y_{ij} = \mu + \beta_i + \varepsilon_{ij}, \tag{1}$$

with mutually independent errors, $\varepsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. We assume that the laboratory random effects $\beta_i \sim N(0, \sigma_B^2)$ are mutually independent and independent with all ε_{ij}. The variance components σ_i^2 and σ_B^2 are the nuisance parameters: the within-laboratory and between-laboratory variances. The parameter of interest is the (unknown) common mean μ. We will use the following notation: $\bar{y}_i = (1/n_i)\sum_{j=1}^{n_i} y_{ij}$, $s_i^2 = (1/(n_i-1))\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$.

The results of typical interlaboratory studies are presented in Table 1 and Table 2, see also e.g. Iyer *et al.* (2002) and Rukhin and Vangel (1998).

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

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

If the variance components σ_B^2 and σ_i^2 would be known the optimal estimator for the unknown common mean μ would be the generalized least squares estimator (GLS estimator) which is under given assumptions MVUE — minimum variance unbiased estimator, that is

$$\hat{\mu}_{GLS} = \frac{\sum_{i=1}^{k} w_i \bar{y}_i}{\sum_{i=1}^{k} w_i}, \tag{2}$$

where $w_i = 1/\mathit{Var}(\bar{y}_i)$ with $\mathit{Var}(\bar{y}_i) = \sigma_B^2 + \sigma_i^2/n_i$. The exact distribution of the estimator is known: $\hat{\mu}_{GLS} \sim N(\mu, 1/\sum w_i)$. From that the standard statistical inference on μ could be performed.

If the variance components are unknown the feasible estimator, typically considered in applications, is that given in (2) with the unknown variance components (consistently) estimated. However, it was demonstrated that the variance of the estimator of the common mean is underestimated by the plug-in statistic leading to too narrow confidence intervals.

The problem considered here is not new and the literature goes back to Cochran (1937). Currently, the problem was studied in details by Rukhin and Vangel (1998), Rukhin, Biggerstaff, and Vangel (2000), Iyer, Wang, and Mathew (2002), and Witkovský and Wimmer (2003).

In this paper we suggest to use a general purpose method for small sample inference for fixed effects in mixed effects models based on restricted maximum likelihood (REML) suggested by Kenward and Roger (1997) and based on the results by Harville and his co-workers, see e.g. Kackar and Harville (1984), Harville and Jeske (1992). Kenward and Roger proposed to use an adjusted F-statistic (in our problem it reduces to the t-statistic) that reduces small sample bias. Its distribution is approximated by F-distribution (t-distribution) with estimated approximate denominator degrees of freedom (DDF). The method could be easily implemented following the calculations detailed by Kenward and Roger (1997). The method was implemented in commercial statistical packages, and hence, it is widely available for applications.

The statistical properties (the empirical coverage probabilities) of the *Kenward-Roger confidence intervals* derived based on large simulation study could be compared with the properties of the other known methods, like e.g. the method suggested in Iyer *et al.* (2002) or in Rukhin and Vangel (1998).

2. Simulation study

In our simulation study we have examined the coverage probabilities of the Kenward-Roger confidence intervals for the common mean μ in model (1). The design of the simulation study is similar to that

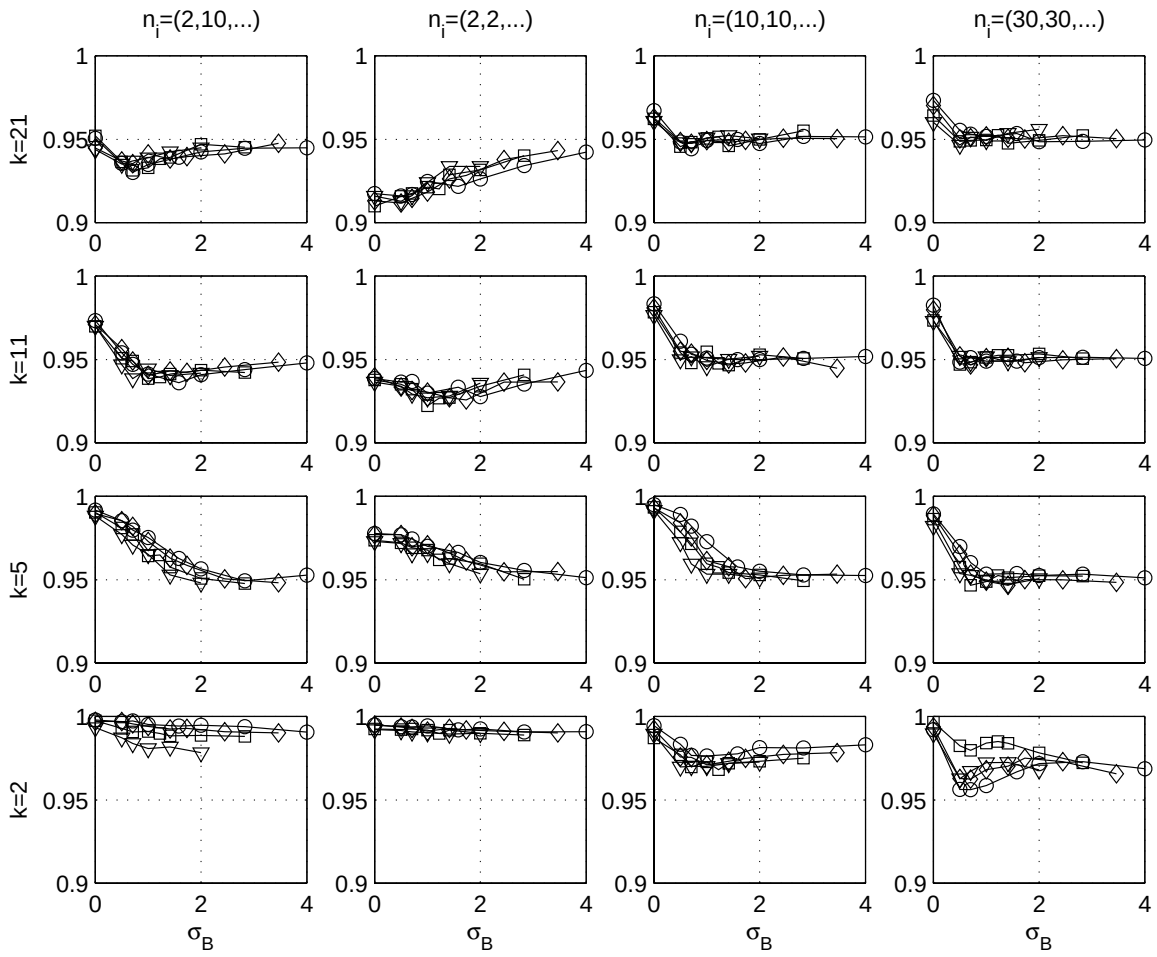

Figure 1: The empirical coverage probabilities of the 95% Kenward-Roger confidence intervals based on 10,000 Monte Carlo runs for each specific design. Here we use the symbol $\bigtriangledown$ for designs with $\sigma_k^2 = 1$, $\square$ for designs with $\sigma_k^2 = 2$, $\diamond$ for designs with $\sigma_k^2 = 3$, and $\bigcirc$ for designs with $\sigma_k^2 = 4$.

presented in Iyer *et al.* (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% Kenward-Roger confidence interval for μ was calculated. The relative frequency of cases such that the Kenward-Roger confidence interval contained the true value $\mu = 0$ was recorded and plotted in Figure 1.

The presented simulation study was performed by using our own MATLAB algorithm, 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). In the case when the calculated denominator degrees of freedom, the DDF, was less than 1 we have used DDF=1. Further, if REML

estimate of any variance component was close to zero, i.e. $\hat{\sigma}^2 < 10^{-6}$, we have changed the elements of the corresponding row and column of the covariance matrix of the REML variance components estimators to zeros.

3. Examples

In Table 3 we present the 95% Kenward-Roger confidence intervals for the common mean of the Selenium in Non-fat Milk Powder Data, see Table 1, and for the common mean of the Arsenic in Oyster Tissue Data, see Table 2, calculated by the MATLAB algorithm. The REML estimates of variance components are: $\sigma_B^2 = 0.0000$, $\sigma_i^2 = \{96.1276, 19.2132, 2.7051, 43.0776\}$ for Selenium in Non-fat Milk Powder Data and $\sigma_B^2 = 1.9142$ and $\sigma_i^2 = \{0.0911, 0.2160, 0.0016, 0.6114, 6.5210, 0.6830, 0.3454, 0.0839, 0.2687, 0.7274, 0.1841, 0.0256, 0.0441, 0.1674, 0.2197, 0.0036, 0.1292, 0.3693, 0.1087, 0.0783, 0.0225, 0.6332, 0.7675, 0.1849, 0.1089, 0.5079, 0.0324, 2.8683\}$ for the Arsenic in Oyster Tissue Data.

Table 3. 95% Kenward-Roger Confidence Intervals for Common Mean

Data	Mean	Std Error	DDF	Lower	Upper
Selenium	109.5788	0.4244	17.5252	108.6854	110.4722
Arsenic	13.2237	0.2686	26.7840	12.6723	13.7752

4. Discussion

In this paper we have suggested to use the *Kenward-Roger confidence intervals* for the common mean in interlaboratory trials. The presented simulation study shows that the suggested confidence intervals have good coverage properties whenever the number of observations in one laboratory is sufficiently large (say 10 observations per one laboratory). The Kenward-Roger method has comparable properties with the more sophisticated method based on the generalized p-values suggested by Iyer *et al.* (2002). Moreover, the Kenward-Roger method has been implemented into the standard statistical software packages like e.g. SAS. The MATLAB algorithm and its R (S-plus) version is available on request from the authors.

References

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

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

[3] Iyer, H.K., Wang, C.M., and Mathew, T. (2002). Models and confidence intervals for true values in interlaboratory trials. *Submitted to the Journal of the American Statistical Association*.

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

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

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

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

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