Available online at www.sciencedirect.com

ScienceDirect

Journal of Statistical Planning and Inference 137 (2007) 264–278

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Statistical aspects of linkage analysis in interlaboratory studies

Andrew L. Rukhin[a, b, *], William E. Strawderman[b, c]

[a]Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA
[b]Statistical Engineering Division, National Institute of Standards and Technology, 820 W. Diamond Ave, Gaithersburg, MD 20899, USA
[c]Department of Statistics, Rutgers University, New Brunswick, NJ 08903, USA

Received 12 July 2004; accepted 14 September 2005
Available online 9 November 2005

Abstract

This paper investigates statistical issues that arise in interlaboratory studies known as Key Comparisons when one has to link several comparisons to or through existing studies. An approach to the analysis of such a data is proposed using Gaussian distributions with heterogeneous variances. We develop conditions for the set of sufficient statistics to be complete and for the uniqueness of uniformly minimum variance unbiased estimators (UMVUE) of the contrast parametric functions. New procedures are derived for estimating these functions with estimates of their uncertainty. These estimates lead to associated confidence intervals for the laboratories (or studies) contrasts. Several examples demonstrate statistical inference for contrasts based on linkage through the pilot laboratories. Monte Carlo simulation results on performance of approximate confidence intervals are also reported.

MSC: Primary 62C15; secondary 62H12; 62F11

Keywords: Behrens–Fisher distribution; Common mean problem; Completeness; Confidence intervals; Contrasts; Graybill–Deal estimator; Key comparisons; Sufficient statistics; Unbiasedness; UMVUE; Welch–Satterthwaite's formula

1. Introduction and summary

The Mutual Recognition Arrangement (MRA) (1999) for national measurement standards is a principal feature of international cooperation for measurement quality assurance. The MRA is realized through Key Comparisons (KC) which typically involve several laboratories with several of them (typically National Metrology Institutes), serving as the *pilot* laboratories designed to coordinate the whole study. Each of the regional laboratories analyzes its measurements and reports the results consisting of its estimate of the measurement value along with the combined standard uncertainty. It is expected that the decomposition of this uncertainty into type A and type B components is presented as required by International Organization for Standardization, Guide to the Expression of Uncertainty in Measurement, (ISO GUM) (ISO, 1993) and NIST Guidelines for Evaluating and Expressing Uncertainty (Taylor and Kuyatt, 1994). The key comparison reference value (KCRV) and its associated uncertainty are determined on the basis of these characteristics. One of the goals is to establish the degree of equivalence of measurements made by participating laboratories and to quantify these degrees of equivalence for all pairs of the laboratories.

* Corresponding author. Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Tel.: +1 41 04552408; fax: +1 41 04551066.
E-mail address: rukhin@math.umbc.edu (A.L. Rukhin).

0378-3758/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jspi.2005.09.004

This paper is motivated by the need for rigorous statistical analysis in this situation where one has to link several comparisons to or through existing KC. In these studies laboratories commonly use a transfer instrument to assess the value of a laboratory standard and to compare the relative biases of their measurement processes and standards. Two important parts of such comparisons are estimates of the difference between two artifacts or between two laboratory effects, i.e., the mentioned degrees of equivalence. There are situations where a direct comparison is not possible because the laboratories have not participated in the same KC study or have not measured the same artifact. In the simplest case one has to seek linking laboratories which have made measurements on the artifacts common to these two laboratories. In this paper we call such laboratories pilot labs.

The MRA does not specify exactly how to perform the linkage. Most recent proposals (e.g. Elster et al., 2003) treat the statistical estimates of uncertainties as known constants which could lead to artificially small confidence intervals for contrasts.

A motivating example for this work is the study by Delahaye and Witt (2002) which was designed to link two existing Capacitance Standards Key Comparisons: CCEM-K4 and the EUROMET project 345 (10 pF results). These two interlaboratory studies were carried out by the Consultative Committee for Electricity and Magnetism (CCEM) and by the European Metrology Cooperation (EUROMET) in the late nineties. Six national institutes (BIPM, CSIRO-NML, NIST, NMi, NPL and PTB) agreed to serve as linking (pilot) labs. These institutes participated in both Key Comparisons, while ten (regional) institutes (BEV, CEM, CMI, CSIR, GUM, IEN, METAS, MIKES/VTT, SP, UME) were additional members of the EUROMET project 345. Delahaye and Witt (2002) formulated the goal of the study as to evaluate "the correction" to the measurements of the labs participating only in the EUROMET project 345 to obtain "the best estimate of what would have been the result from such a laboratory had it actually participated in CCEM-K4". After this correction has been found, the table of pairwise laboratory contrasts or bilateral equivalences along with associated (combined) uncertainties is determined.

The main contribution of this paper is a procedure implementing such a correction along with the characteristic uncertainty of the resulting estimates. In Section 2 we suggest a linear model for linkage. Conditions for completeness of sufficient statistics are derived and applied to an example in Section 3. Several models for type B uncertainty are presented and discussed in Section 4. The situation when the sufficient statistics are not complete is investigated in Section 5. A curious and potentially useful aspect of such non-saturated designs is that there may exist UMVUE's for certain contrasts of interest. In Section 6 confidence intervals for contrasts are discussed and compared, and some Monte Carlo simulation results for the coverage probability of the developed approximate confidence intervals are reported. The paper is concluded with discussion in Section 7 and the Appendix which contains selected proofs.

2. The basic model

We use a natural model, which assumes that a number, say, K of laboratories serve as pilot labs measuring several of J given artifacts (or participating in several out of J different studies.) Non-pilot laboratories measure only one of these artifacts with data in each laboratory having an additive error structure. More precisely, we investigate the following linear model for the data Y_{ijk}:

$$Y_{ijk} = \ell_i + a_j + e_{ijk}.$$

Here $i = 1, \ldots, M$ indexes the laboratories, $j = 1, \ldots, J$ corresponds to different artifacts, and $n_{ij} \geqslant 2$, represents the sample size (the number of measurements) in laboratory i which measures artifact j; ℓ_i is the effect of the ith laboratory and a_j is the effect of jth artifact. It will be assumed that $e_{ijk} \sim \mathrm{N}(0, \tau_{ij}^2)$, $k = 1, \ldots, n_{ij}$ are mutually independent. Notice that the variances τ_{ij}^2 are not assumed to be equal as in standard linear normal models. However, unequal error variances are common, indeed virtually universal, in interlaboratory studies. In some situations it is assumed that the variances τ_{ij} do not depend on j. We do not make this assumption; all results with minor modifications hold in the model when the variances do not depend on the artifact measured.

This model assumes that the results of all measurements are of the same order of magnitude and of the same physical dimension. If this condition is not met, a transformation of variables of the type suggested by Elster et al. (2003) may be necessary.

The vector of sufficient statistics for unknown parameters $\ell_i + a_j$ and τ_{ij}^2 is formed by the sample means $Y_{ij} = \sum_k Y_{ijk}/n_{ij}$, and by sample variances $v_{ij}^2 = \sum_k (Y_{ijk} - Y_{ij})^2/\kappa_{ij}$ with $\kappa_{ij} = n_{ij} - 1$. Clearly,

$$Y_{ij} = \ell_i + a_j + e_{ij} \tag{1}$$

with independent $e_{ij} \sim \mathrm{N}(0, \sigma_{ij}^2 = \tau_{ij}^2/n_{ij})$ and $\kappa_{ij} v_{ij}^2/\tau_{ij}^2 \sim \chi^2_{\kappa_{ij}}$. It is convenient to put $n = \sum_{ij} n_{ij}$, and $s_{ij}^2 = v_{ij}^2/n_{ij}$ as this is a classical estimate of the variance σ_{ij}^2 of the sample mean Y_{ij}. This and also the unequal variances condition make application of standard results from two-way classified data designs infeasible.

Let G_j denote the set of all laboratories which measure artifact j, $j = 1, \ldots, J$. Then for $j \neq j'$, $G_j \cap G_{j'}$ consists of all pilot laboratories measuring both artifacts j and j' (or is the empty set.) Further, we use notation H_i for the set of artifacts measured by the lab i.

Observe that in our model the individual laboratory effects ℓ_i and artifact effects a_j are not identifiable (as the sum $\ell_i + a_j$ is unchanged by adding a constant to all a's and subtracting a constant from all ℓ's). However, any contrast in $\ell_i, i \in G_j$, measuring the same artifact or any linear combination of such contrasts is estimable. Indeed, our main problem of interest is to estimate the difference $a_j - a_{j'}$ between two artifacts, or the contrast $\ell_i - \ell_{i'}$ for two laboratories i and i'. We seek estimators with the smallest mean squared error.

If L_i and $L_{i'}$ are two laboratories, we write $L_i \sim L_{i'}$, if they measure the same artifact. One can form a graph **G** with vertexes formed by all laboratories, and edges connecting any two vertexes measuring the same artifact, i.e. if they are equivalent under relation "$\sim$".

Assumption 1. The graph **G** is connected, i.e. for any two laboratories L_i and $L_{i'}$ there exists a sequence of labs (path) $L_{i_1}, \ldots, L_{i_m}$ such that $L_i \sim L_{i_1} \sim \cdots \sim L_{i_m} \sim L_{i'}$.

Notice that in this graph the non-pilot labs are connected to the pilot labs measuring the same artifact. Then n, the total number of labs, is the total number of vertexes in the graph **G**.

The motivation for our setting comes from the existing Key Comparison projects organized by CCEM, in particular, from the Key Comparisons Capacitance Standards linkage problem introduced in Section 1. In this example, artifacts correspond to different studies (say, a_1 means the EUROMET project 345, and a_2 corresponds to CCEM-K4), so that $J = 2$, and $K = 6$. In this study G_2 had four non-pilot laboratories (MSL, NIM, NRC and VNIM), so that $n = 20$; $G_1 \cap G_2$ consists of mentioned six pilot laboratories, and, in addition, G_1 has 10 regional institutes listed in Section 1. The parameter of interest is the difference $a_1 - a_2$ between two studies, since by subtracting this term form the measurements $Y_{i1}, i \in G_1$, one indeed obtains "the best estimate of what would have been the result from such a laboratory had it actually participated in CCEM-K4". Clearly, a good estimate of $a_1 - a_2$ leads to an estimate of $\ell_i - \ell_{i'}$ for laboratory $i, i \in G_1$, and $i', i' \in G_2$ (and reversely).

In another, more complicated, study CCT-K3 done under the auspices of the Consultative Committee for Thermometry (Magnum et al., 2002) there were 9 different serial numbered thermometers (artifacts) with 3 pilot labs (NIST, NML and PTB) (some of which measured more than 2 of these thermometers) and 12 regional laboratories (most of which measured just 1 thermometer.) The goal here was to obtain the table of differences of realizations between these fifteen laboratories along with their expanded uncertainties. Albeit more involved, such a table can be derived by methods of this paper if one studies linkage problems for labs measuring only a particular subset of all thermometers.

3. Conditions for completeness and confidence intervals for contrasts (saturated models)

The following general result is useful in our study of linkage analysis.

Proposition 1. *Let $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ and let S_i^2 be independent of X_i, so that S_i^2/σ_i^2 has a χ^2-distribution with κ_i degrees of freedom, $i = 1, \ldots, m$. Assume that the vector $(\mu_1, \ldots, \mu_m)$ belongs to a vector space Θ of dimension p, $p \leqslant m$, and the values of σ_i^2 are unrestricted. Then a necessary and sufficient condition for $(X_1, \ldots, X_m, S_1^2, \ldots, S_m^2)$ to form a complete sufficient statistic for unknown parameters $(\mu_1, \ldots, \mu_m, \sigma_1^2, \ldots, \sigma_m^2)$ is that $p = m$.*

The proof of Proposition 1 follows from standard facts about exponential families. Notice that if the vector $c = (c_1, \ldots, c_m)$ is orthogonal to Θ, then the linear combination $\sum c_i X_i$ is an unbiased estimator of zero. According to Proposition 1 when $p = m$, each linear combination of the X's is the UMVUE of its expected value.

We show that under an additional assumption, Y_{ij}'s form a complete sufficient statistic if and only if the number of pilot laboratories is equal to the number of artifacts minus one.

Proposition 2. *Assume that each pilot laboratory is in exactly two of G_j's and each non-pilot lab belongs exactly to one such set. Then under Assumption 1 the vector $Y_{ij}, s_{ij}^2, i \in G_j, j = 1, \ldots, J$, forms a complete sufficient statistic if and only if $J = K + 1$.*

Proposition 2 whose proof is in the Appendix gives conditions under which a design has complete sufficient statistics, i.e., is saturated. In the most straightforward case, the first pilot lab measures artifacts a_1 and a_2, the second pilot lab measures a_2 and a_3, etc, the last Kth pilot lab measures a_K and a_{K+1}. But, for example, the design where all pilot labs measure artifact a_{K+1}, while the ith laboratory measures artifact $a_i, i = 1, \ldots, K$, is also saturated in this sense.

Whenever two different pilot labs (say L_1 and L_2) measure the same two artifacts, say, a_1 and a_2, completeness is lost, as

$$E \sum_{i=1}^{2} \sum_{j=1}^{2} (-1)^{i+j} Y_{ij} = \sum_{i=1}^{2} \sum_{j=1}^{2} (-1)^{i+j} (\ell_i + a_j) = 0.$$

This also is the case for more complicated schemes in which the same two (or more) artifacts circulate through the same two (or more) labs. We study non-saturated designs in Section 5.

Proposition 3. *Under conditions of Proposition 2, when $K = J - 1$, any estimator*

$$\delta = \sum_j \sum_{i \in G_j} b_{ij} Y_{ij} = \sum_i \sum_{j \in H_i} b_{ij} Y_{ij} \tag{2}$$

is a UMVUE of the parametric function $\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j)$.

In particular, if for each j, $\sum_{i \in G_j} b_{ij} = 0$, then $E\delta = \sum_j \sum_{i \in G_j} b_{ij} \ell_i$. Similarly, if for each i, $\sum_{j \in H_i} b_{ij} = 0$, the expected value is $\sum_j \sum_{i \in G_j} b_{ij} a_j$. If $K < J - 1$, this estimator is unbiased, but not in general UMVUE.

Proposition 4. *The variance of estimator (2) has the form*

$$Var(\delta) = \sum_j \sum_{i \in G_j} b_{ij}^2 \sigma_{ij}^2,$$

which can be unbiasedly estimated via $\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2$.

Note that the estimators which are UMVUE of the linear combinations of $\ell_i + a_j$ under the assumption of normality will be best linear unbiased estimators under general distributional assumptions.

As was mentioned in Section 2, any contrast in $\ell_i, i \in G_j$, is unbiasedly estimable. In particular, all comparisons of the form $\ell_i - \ell_{i'}$ and $a_j - a_k$ are estimable. These unbiased estimators are obtained from the linkage through the intermediate pilot labs. We start with the approximate Welch–Satterthwaite confidence interval (Welch, 1937), the use of which is recommended by the ISO. The following example illustrates this process.

Example 1. Suppose that there are three artifacts, a_1, a_2 and a_3, the first of which circulates through labs $L_i, i = 1, \ldots, I$. The second artifact is measured by labs with $i = I, \ldots, M$, and the last one circulates through labs $L_i, i = M, \ldots, n (I < M < n)$. Thus, there are two pilot labs L_I and L_M. We start with the interlaboratory difference, $\ell_i - \ell_{i'}$, which admits a UMVUE according to Proposition 2.

1. If $1 \leqslant i, i' \leqslant I$, the UMVUE is $Y_{i1} - Y_{i'1}$, with the variance $\sigma_{i1}^2 + \sigma_{i'1}^2$. An unbiased estimator of this variance has the form $s_{i1}^2 + s_{i'1}^2$. An approximate Welch–Satterthwaite confidence interval for $\ell_i - \ell_{i'}$ is given by $Y_{i1} - Y_{i'1} \pm t\sqrt{s_{i1}^2 + s_{i'1}^2}$, where t is the upper $\alpha/2$ point of a t-distribution with $v = (s_{i1}^2 + s_{i'1}^2)^2 / (s_{i1}^4/\kappa_{i1} + s_{i'1}^4/\kappa_{i'1})$ degrees of freedom.

2. If $1 \leqslant i \leqslant I-1$, $I+1 \leqslant i' \leqslant M$, the UMVUE is $Y_{i1} - Y_{I1} + Y_{I2} - Y_{i'2}$, so that this estimate is "linked" through the pilot lab L_I. The variance now is $\sigma_{i1}^2 + \sigma_{I1}^2 + \sigma_{I2}^2 + \sigma_{i'2}^2$ with an unbiased estimator $s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{i'2}^2$. Here the requirement that we link through the pilot lab L_I increases the variance of the estimated contrast. The degrees of freedom for the approximate Welch–Satterthwaite confidence interval are

$$\nu = \frac{(s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{i'2}^2)^2}{s_{i1}^4/\kappa_{i1} + s_{I1}^4/\kappa_{I1} + s_{I2}^4/\kappa_{I2} + s_{i'2}^4/\kappa_{i'2}}.$$

3. If $1 \leqslant i \leqslant I-1$, $M+1 \leqslant i' \leqslant N$, i.e. if lab i measured artifact 1, and lab i' measured the artifact 3, the UMVUE is $Y_{i1} - Y_{I1} + Y_{I2} - Y_{M2} + Y_{M3} - Y_{i'3}$, so that this estimate is linked through the pilot labs L_I and L_M. The variance of this estimator is $\sigma_{i1}^2 + \sigma_{I1}^2 + \sigma_{I2}^2 + \sigma_{M2}^2 + \sigma_{M3}^2 + \sigma_{i'3}^2$ with an unbiased estimator $s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{M2}^2 + s_{M3}^2 + s_{i'3}^2$. Thus, the additional linkage through the pilot lab L_M further increases the variance. The Welch–Satterthwaite confidence interval is based on

$$\nu = [(s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{M2}^2 + s_{M3}^2 + s_{i'3}^2)^2]\left[\frac{s_{i1}^4}{\kappa_{i1}} + \frac{s_{I1}^4}{\kappa_{I1}} + \frac{s_{I2}^4}{\kappa_{I2}} + \frac{s_{M2}^4}{\kappa_{M2}} + \frac{s_{M3}^4}{\kappa_{M3}} + \frac{s_{i'3}^4}{\kappa_{i'3}}\right]^{-1}$$

degrees of freedom.

The unbiased estimator of the difference $a_1 - a_2$ is based on $Y_{I1} - Y_{I2}$, of $a_2 - a_3$ on $Y_{M2} - Y_{M3}$, and to estimate $a_1 - a_3$ one has to use $Y_{I1} + Y_{M2} - Y_{I2} - Y_{M3}$. The degrees of freedom for the confidence interval are $(s_{I1}^2 + s_{I2}^2)^2/(s_{I1}^4/\kappa_{I1} + s_{I2}^4/\kappa_{I2})$, $(s_{M2}^2 + s_{M3}^2)^2/(s_{M2}^4/\kappa_{M2} + s_{M3}^4/\kappa_{M3})$, and $(s_{I1}^2 + s_{I2}^2 + s_{M2}^2 + s_{M3}^2)^2/(s_{I1}^4/\kappa_{I1} + s_{I2}^4/\kappa_{I2} + s_{M2}^4/\kappa_{M2} + s_{M3}^4/\kappa_{M3})$, respectively.

More generally, the approximate Welch–Satterthwaite confidence interval for the function $\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j)$ is $\delta \pm t\sqrt{\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2}$, where t denotes the critical $\alpha/2$ point of a t-distribution with

$$\nu = \frac{\left(\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2\right)^2}{\sum_j \sum_{i \in G_j} b_{ij}^4 s_{ij}^4/\kappa_{ij}} \tag{3}$$

degrees of freedom. We consider an alternative approximation by a multiple of a t-random variable for non-saturated designs in Section 6.

When $n = 2$, $\ell_1 = \ell_2 = 0$, the situation is that of the Behrens–Fisher problem, where a confidence interval for the difference between two normal means $a_1 - a_2$ is desired. This shows that only approximate confidence intervals for the contrasts can be anticipated as similar tests or confidence intervals do not exist in this problem (Linnik, 1968).

4. Type B error

It has become common in interlaboratory studies to include a Type B error in the final analysis. Typically, the Type A error is the standard experimental error captured by the residual term in the model, and it can be measured by the estimated variances s_{ij}^2 (or by some other statistical procedures.) The type B error, on the other hand, cannot be measured through experimental replication, but rather through a process of creating an uncertainty budget for all important sources of variability that could have affected the experimental result. This process results in a total type B uncertainty which must be included in error assessments of all estimates and decisions.

We start with a fairly simple model that embodies the type B error as follows. The sample means now have the form

$$Y_{ij} = \ell_i + a_j + b_i + e_{ij} \tag{4}$$

for $i \in G_j$, $j = 1, \ldots, J$. Here ℓ_i, a_j, e_{ij} have the same meaning as in (1); b_i corresponds to independent realizations of normal $\mathrm{N}(0, \beta_i^2)$ random variables independent of e_{ij} and of the estimates s_{ij}^2. The values β_i^2 are assumed to be known and taken to be the Type B error specified in the uncertainty budget. Model (4) falls into domain of variance components analysis, but in our case one of the variances is known.

If the mean b_i is non-zero, but is known, then, according to ISO GUM (1993) recommendations, it must be subtracted from the data. When it is unknown, the problem is much more difficult. Model (4) assumes an accurate uncertainty budget. Even in highly qualified national institutes the possibility of inaccurate budgets are real, and we attempt to allow for this later in this section.

The following result is an extension of Proposition 1 to cover the present model. Its proof is contained in the Appendix.

Proposition 5. *Under conditions of Proposition* 1 *suppose that a random vector* $B \sim \mathrm{N}_m(0, V)$ *with possibly singular matrix* V *is independent of the vector* X *with independent coordinates* $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ *and of* S_i^2, *with* S_i^2/σ_i^2 *having a* χ^2*-distribution. Assume that the vector* $(\mu_1, \ldots, \mu_m)$ *belongs to a vector space* Θ *of dimension* p, $p \leqslant m$. *Then if the observed data are* (Y, S^2) *with* $Y = X + B$, $S^2 = (S_1^2, \ldots, S_m^2)$, *a necessary and sufficient condition for* (Y, S^2), *to form a complete sufficient statistic for unknown parameters* $(\mu_1, \ldots, \mu_m, \sigma_1^2, \ldots, \sigma_m^2)$ *is that* $p = m$.

The next result is an application of Propositions 4 and 5.

Proposition 6. *Estimator* (2) *remains unbiased in model* (4). *Under conditions of Proposition* 2 *it is UMVUE if* $K = J - 1$. *Its variance has the form*

$$Var(\delta) = \sum_j \sum_{i \in G_j} b_{ij}^2 \sigma_{ij}^2 + \sum_i \left(\sum_{j \in H_i} b_{ij} \right)^2 \beta_i^2. \tag{5}$$

An unbiased estimator of this variance has the form

$$\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2 + \sum_i \left(\sum_{j \in H_i} b_{ij} \right)^2 \beta_i^2. \tag{6}$$

The formula for the variance follows since $EY_{ij}Y_{k\ell} = \delta_{ik}\delta_{j\ell}\sigma_{ij}^2 + \delta_{ik}\beta_i^2$. Here and further δ_{ik} is the Kronecker symbol.

The approximate Welch–Satterthwaite confidence interval for the function $\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j)$ is

$$\delta \pm t \sqrt{\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2 + \sum_i \left(\sum_{j \in H_i} b_{ij} \right)^2 \beta_i^2},$$

where t is the upper $\alpha/2$ point of a t-distribution with

$$v = \frac{\left[\sum_j \sum_{i \in G_j} b_{ij}^2 s_{ij}^2 + \sum_i \left(\sum_{j \in H_j} b_{ij} \right)^2 \beta_i^2 \right]^2}{\sum_j \sum_{i \in G_j} b_{ij}^4 s_{ij}^4 / \kappa_{ij}}$$

degrees of freedom.

Let us return to Example 1 of Section 4 with Type B error. The data are still complete and sufficient because of Proposition 4, and the estimators of $a_1 - a_2$ or of $\ell_i - \ell_{i'}$ are UMVUE's by completeness.

1. If $1 \leqslant i, i' \leqslant I$, the UMVUE of $\ell_i - \ell_{i'}$ is $Y_{i1} - Y_{i'1}$, with the variance $\sigma_{i1}^2 + \sigma_{i'1}^2 + \beta_i^2 + \beta_{i'}^2$. An unbiased estimator of this variance has the form, $s_{i1}^2 + s_{i'1}^2 + \beta_i^2 + \beta_{i'}^2$. An approximate Welch–Satterthwaite confidence interval has the form $Y_{i1} - Y_{i'1} \pm t\sqrt{s_{i1}^2 + s_{i'1}^2 + \beta_i^2 + \beta_{i'}^2}$, where t is the upper $\alpha/2$ point of a t-distribution with $v = (s_{i1}^2 + s_{i'1}^2 + \beta_i^2 + \beta_{i'}^2)^2/(s_{i1}^4/\kappa_{i1} + s_{i'1}^4/\kappa_{i'1})$ degrees of freedom.

2. If $1 \leqslant i \leqslant I-1$, $I+1 \leqslant i' \leqslant M$, the UMVUE still is $Y_{i1} - Y_{I1} + Y_{I2} - Y_{i'2}$. Its variance is $\sigma_{i1}^2 + \sigma_{I1}^2 + \sigma_{I2}^2 + \sigma_{i'2}^2 + \beta_i^2 + \beta_{i'}^2$. An unbiased estimator is $s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{i'2}^2 + \beta_i^2 + \beta_{i'}^2$. The degrees of freedom for the approximate Welch–Satterthwaite

confidence interval are

$$v = \frac{(s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{i'2}^2 + \beta_i^2 + \beta_{i'}^2)^2}{s_{i1}^4/\kappa_{i1} + s_{I1}^4/\kappa_{I1} + s_{I2}^4/\kappa_{I2} + s_{i'2}^4/\kappa_{i'2}}.$$

3. If $1 \leqslant i \leqslant I-1$, $M+1 \leqslant i' \leqslant N$, the UMVUE is $Y_{i1} - Y_{I1} + Y_{I2} - Y_{M2} + Y_{M3} - Y_{i'3}$, and the variance of this estimator is $\sigma_{i1}^2 + \sigma_{I1}^2 + \sigma_{I2}^2 + \sigma_{M2}^2 + \sigma_{M3}^2 + \sigma_{i'3}^2 + \beta_i^2 + \beta_{i'}^2$ with an unbiased estimator $s_{i1}^2 + s_{I1}^2 + s_{I2}^2 + s_{M2}^2 + s_{M3}^2 + s_{i'3}^2 + \beta_i^2 + \beta_{i'}^2$. The degrees of freedom of the Welch–Satterthwaite confidence interval are determined in a similar way.

Note that the type B errors associated with the linked pilot labs play no role in the variances of the lab comparisons in Example 1. However, they do enter into the variances of the estimators $Y_{I1} - Y_{I2}$, $Y_{M2} - Y_{M3}$, or $Y_{I1} + Y_{M2} - Y_{I2} - Y_{M3}$ of the artifact comparisons $a_1 - a_2$, $a_2 - a_3$, or $a_1 - a_3$.

Often a Type B uncertainty is expressed as a best estimate of the variance of the Type B error together with associated degrees of freedom. We suggest the following models to describe this situation. In each of them, as in (4),

$$Y_{ij} = \ell_i + a_j + B_i + e_{ij} \tag{7}$$

for $i \in G_j$, $j = 1, \ldots, J$. Here ℓ_i, a_j, e_{ij} have the same meaning as in (1) and (4), but B_i corresponds to independent realizations of random variables with mean zero, also independent of e_{ij} and of the estimates s_{ij}^2. However, the variances of these variables, β_i^2, are not assumed to be known. Assume first that a realization of a random variable $\rho_i^2 \sim \beta_i^2 \chi_{v(i)}^2 / v(i)$, with given $v(i)$, is available. Then the Type B uncertainty is expressed through the pair $(\rho_i^2, v(i))$ which is arrived at through the uncertainty budget analysis. In this sense Type B error assessment is uncertain, and the case where $v(i) = \infty$ corresponds to the previous setting.

In the second interpretation of (7), β_i^2 is random (so that B_i is normal only conditionally) with the distribution of the form $\rho_i^2 \chi_{v(i)}^2 / v(i)$ with known ρ_i^2 and $v(i)$. The Type B error is still expressed through the pair $(\rho_i^2, v(i))$.

Proposition 6 remains valid for this model with essentially the same proof except that in the second case in (5) the unknown β_i^2 are to be replaced by the known ρ_i^2. Besides that, in the first case the modification to be made in Example 1 is that in the formula for the degrees of freedom v, the denominator must have an additional term $\rho_i^4/v(i) + \rho_{i'}^4/v(i')$, and in (6), β_i^2 are to be replaced by ρ_i^2.

Still another option under model (1) is to interpret Type B uncertainty for lab i as the parameter(s) of the prior inverse gamma-density,

$$\pi_i(v) = \frac{\exp\{-1/(\lambda_i v)\}}{\lambda_i^{\alpha_i} \Gamma(\alpha_i) v^{\alpha_i + 1}}, \tag{8}$$

$v = \sigma_{ij}^2$ (assuming that this distribution does not depend on j.) More precisely, the effective degree of freedom $2\alpha_i$ in (8) can be taken to be $v(i)$, whereas the scale parameter λ_i is inversely proportional to uncertainty ρ_i^2, $\lambda_i = 1/[\rho_i^2(0.5v(i) - 1)]$. This formula is a consequence of the equation, $E\sigma_{ij}^2 = 1/[\lambda_i(\alpha_i - 1)]$. This model leads to usable formulas as for fixed s_{ij}^2,

$$Var(Y_{ij}) = \frac{\kappa_{ij} s_{ij}^2 + 0.5/\lambda_i}{n_{ij}(\kappa_{ij} + 2\alpha_i - 2)},$$

and a (conditional) confidence interval for any linear combination $\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j)$ can be obtained from a t-approximation as elaborated in Section 6.

5. UMVUE when sufficient statistics are incomplete and unbiased estimators for non-saturated models

It may still happen that certain estimators are UMVUE's of their expectations even when the linkage design is not saturated. In particular, the following result is useful in the context of Proposition 1 (cf. Lehmann, 1983 Lemma 3.2, which is a similar result for admissible estimators). Its proof as well as that of the following Proposition 8 are given in the Appendix.

Proposition 7. *Suppose X has a distribution P_θ for $\theta \in \Theta$ and Y has a distribution P_η for $\eta \in H$, and X and Y are independent random variables in the model with the parameter space $\Theta \times H$. If $\delta(X)$ is UMVUE of $g(\theta)$, when only X is observed. $\delta(X)$ remains UMVUE when (X, Y) is observed.*

Proposition 8. *In the model of Proposition 1, suppose that $p < m$, but the $m - p$ independent linear restrictions on $(\mu_1, \ldots, \mu_m)$ do not involve $(\mu_1, \ldots, \mu_r)$, $r < m$. Then any estimator $\delta(X_1, \ldots, X_r, S_1^2, \ldots, S_r^2)$ is a UMVUE of its expectation.*

An implication of this result is that $Y_{ij} - Y_{i'j}$ is always a UMVUE of the contrast $\ell_i - \ell_{i'}$ if i and i' correspond to two non-pilot labs within one group G_j. Similarly, if, say, lab i is in G_1 and lab i' is in G_2, $J = 3$, $H_1 = \{a_1, a_2\}$, $H_2 = \{a_2, a_3\}$, $H_3 = \{a_1, a_3\}$, i.e., G_1 and G_2 have a single pilot lab, say L_1, in common, and G_2 has two other pilot labs, say, L_2 and L_3, which both also measure artifact a_3, then the estimator $Y_{i1} - Y_{11} + Y_{12} - Y_{i'2}$ is UMVUE of $\ell_i - \ell_{i'}$. This follows from Proposition 8 because $Y_{i1}, Y_{i'2}, Y_{11}, Y_{12}, s_{i1}^2, s_{i'2}^2, s_{11}^2, s_{12}^2$ form a complete sufficient subset of the full data set, and the means of these Y's are not involved in any linear restrictions.

Also, if a pilot lab L_I is the only one such lab measuring artifacts a_j and $a_{j'}$, i.e., if $G_j \cap G_{j'}$ has a unique pilot lab L_I, then the UMVUE of the difference $a_j - a_{j'}$ is $Y_{Ij} - Y_{Ij'}$. The fact that the estimators of the variances of these estimators are also UMVUE's follows from Proposition 8.

In general, there will be no UMVUE of $a_j - a_{j'}$ or $\ell_i - \ell_{i'}$ if the comparison made between two labs must be linked through a pilot lab which is involved in the constraints in Proposition 8, and there will be no UMVUE for the artifacts comparisons if these are measured by two (or more) different pilot labs. In such cases a reasonable approach is to find an estimator expressed as the weighted linear combination of Y_{ij} where the weights depend on the sample variances. Assume that $i \in G_1$, $i' \in G_2$, and we are interested in the contrast $a_1 - a_2$ or $\ell_i - \ell_{i'}$.

Example 2. Let us start with the case when G_1 and G_2 have at least two pilot labs in common. To be specific suppose that with $J = 2$, $H_1 = H_2 = \{a_1, a_2\}$, so that pilot labs L_1 and L_2 are in $G_1 \cap G_2$.

Assume that neither of the two groups has any other pilot labs. In this case there is no UMVUE of $a_1 - a_2$ or $\ell_i - \ell_{i'}$. The estimator

$$Y_{i1} - w(Y_{11} - Y_{12}) - (1 - w)(Y_{21} - Y_{22}) - Y_{i'2}$$

is unbiased for $\ell_i - \ell_{i'}$ for any real w. Under model (4), the variance of this estimator is

$$\sigma_{i1}^2 + \sigma_{i'2}^2 + \beta_i^2 + \beta_{i'}^2 + w^2(\sigma_{11}^2 + \sigma_{12}^2) + (1 - w)^2(\sigma_{21}^2 + \sigma_{22}^2).$$

The minimum value is attained when $w = (\sigma_{21}^2 + \sigma_{22}^2)/(\sigma_{11}^2 + \sigma_{12}^2 + \sigma_{21}^2 + \sigma_{22}^2)$. Note that Type B variances for the pilot labs, β_i^2 and $\beta_{i'}^2$, play no role in the expression for w. It is not hard to see that with this optimal choice, the estimator above would be UMVUE if all variances were known. (Indeed, when all variances are known, there is a complete sufficient statistic whose dimension is that of the space spanned by the mean vector.) But since the optimal weights depend on these variances, there is no UMVUE.

However, unbiasedness holds if the weight w is a function of the variance components as they are independent of Y_{ij}. Provided that the degrees of freedom of s_{ij}^2 are sufficiently large, a reasonable estimator of the optimal weight is $\hat{w} = (s_{21}^2 + s_{22}^2)/(s_{11}^2 + s_{12}^2 + s_{21}^2 + s_{22}^2)$ with the resulting estimator of $a_1 - a_2$,

$$\psi = \hat{w}(Y_{11} - Y_{12}) + (1 - \hat{w})(Y_{21} - Y_{22}), \tag{9}$$

and the estimator, $Y_{i1} - \psi - Y_{i'2}$, of $\ell_i - \ell_{i'}$.

The usual estimator of the variance of ψ in (9) is

$$\widehat{Var(\psi)} = \hat{w}^2(s_{11}^2 + s_{12}^2) + (1 - \hat{w})^2(s_{21}^2 + s_{22}^2).$$

The relationship of this problem to that of a common mean estimation omnipresent in interlaboratory studies is quite clear. Indeed, $Y_{11} - Y_{12}$ and $Y_{21} - Y_{22}$ are two independent unbiased estimators of $a_1 - a_2$. However, here, unlike the usual version of the common mean problem, the variance estimators (of the two independent estimates of the common mean) are not distributed as multiples of χ^2-distributions. Still estimator (9) bears considerable resemblance to the popular Graybill and Deal (1959) estimator. Unbiased estimators of the common mean and of their variance is reviewed in Voinov and Nikulin (1993, pp. 194–196).

A similar analysis can be made to obtain unbiased estimators when the linkage must be made through more than one node. For instance, assume that pilot labs L_1 and L_2 are in the set $G_1 \cap G_2$, which contains other pilot labs.

Example 3. Assume that the pilot labs form a path of the form $L_1 \sim L_2 \sim L_3 \sim L_1$. More specifically, with $J = 3$, $H_1 = \{a_1, a_2\}$, $H_2 = \{a_2, a_3\}$, $H_3 = \{a_1, a_3\}$, and $L_1, L_2 \in G_1 \cap G_2$.

In this situation the linkage can be made "forward" through L_1 or "backward" through L_3 and L_2. The first unbiased estimator of $\ell_i - \ell_{i'}$,

$$\delta_0 = Y_{i1} - Y_{11} + Y_{12} - Y_{i'2},$$

corresponds to the first type of linkage, the second estimator,

$$\delta_1 = Y_{i1} - Y_{31} + Y_{33} - Y_{23} + Y_{22} - Y_{i'2},$$

to the "backward" linkage. A broader class obtains by taking convex linear combinations of these statistics,

$$\delta_w = w\delta_1 + (1 - w)\delta_0 = Y_{i1} - w(Y_{31} - Y_{33} + Y_{23} - Y_{22}) - (1 - w)(Y_{11} - Y_{12}) - Y_{i'2}.$$

If all variances were known, the optimal choice for w is

$$w = \frac{\sigma_{11}^2 + \sigma_{12}^2}{\sigma_{31}^2 + \sigma_{33}^2 + \sigma_{23}^2 + \sigma_{22}^2 + \sigma_{11}^2 + \sigma_{12}^2},$$

which does not involve the β's for the pilot labs.

As in the previous example, it can be shown that this choice of w gives the UMVUE for the known variance case, so that there is no UMVUE when the variance is unknown. However, one can estimate w by $\hat{w}$ obtained by substituting s_{ij}^2 for σ_{ij}^2 above; then

$$\psi = \hat{w}(Y_{31} - Y_{33} + Y_{23} - Y_{22}) + (1 - \hat{w})(Y_{11} - Y_{12})$$

is the weighted means estimator of $a_1 - a_2$ which again plays the role of a common mean.

Example 4. This example illustrates linking through two sets of pilot labs. Assume again that there are three artifacts, $J = 3$, and four pilot labs of which L_1 and L_2 are common to G_1 and G_3, while L_3 and L_4 are common to G_2 and G_3. Thus, $H_1 = H_2 = \{a_1, a_3\}$, $H_3 = H_4 = \{a_2, a_3\}$. In this case $G_1 \cap G_2 = \emptyset$.

Estimators of the form

$$\psi = w(Y_{11} - Y_{13}) + (1 - w)(Y_{21} - Y_{23}) + u(Y_{33} - Y_{32}) + (1 - u)(Y_{43} - Y_{42}) \tag{10}$$

are unbiased for $a_1 - a_2$; the optimal choices of w and u are

$$w = \frac{\sigma_{21}^2 + \sigma_{23}^2}{\sigma_{21}^2 + \sigma_{23}^2 + \sigma_{11}^2 + \sigma_{13}^2}, \qquad u = \frac{\sigma_{43}^2 + \sigma_{42}^2}{\sigma_{43}^2 + \sigma_{42}^2 + \sigma_{33}^2 + \sigma_{32}^2}.$$

If the variances were known, these choices of w and u result in UMVUE, so in the unknown variance case there is no UMVUE. A reasonable unbiased estimator is obtained by substituting the following estimates of w and u,

$$\hat{w} = \frac{s_{21}^2 + s_{23}^2}{s_{21}^2 + s_{23}^2 + s_{11}^2 + s_{13}^2}, \qquad \hat{u} = \frac{s_{43}^2 + s_{42}^2}{s_{43}^2 + s_{42}^2 + s_{33}^2 + s_{32}^2}.$$

In this situation one can think of two non-standard common mean problems, one with $a_1 - a_3$ and another with $a_3 - a_2$, playing the role of a common mean which admits several unbiased independent estimators. The corresponding unbiased estimator for $\ell_i - \ell_{i'}$ is $Y_{i1} - \psi - Y_{i'2}$; the optimal choices of w and u remain the same.

These examples give rise to the following general result. Denote by $\mathscr{P}$ a generic path passing through pilot labs in the graph $\mathbf{G}$, $L_{i_0} \sim L_{i_1} \sim \cdots \sim L_{i_M}$, and such that $L_{i_0} \in G_1 \cap G_{j_1}$, $L_{i_1} \in G_{j_1} \cap G_{j_2}, \ldots, L_{i_M} \in G_{j_M} \cap G_2$, $i_0 \neq i_1 \neq \cdots \neq i_M$. We put $j_0 = 1$, $j_{M+1} = 2$, $\mathbf{e} = (1, \ldots, 1)^{\mathrm{T}}$, and write

$$\mathscr{P} = \begin{pmatrix} i_0 & i_1 & \ldots & i_M \\ 1 & j_1 & \ldots & j_M \end{pmatrix}.$$

Proposition 9. *Under model* (1), *each path $\mathscr{P}$ as above leads to an unbiased estimator of* $a_1 - a_2 = \sum_{k=0}^{M}(a_{j_k} - a_{j_{k+1}})$,

$$\psi_{\mathscr{P}} = \sum_{k=0}^{M} (Y_{i_k j_k} - Y_{i_k j_{k+1}}),$$

and to an unbiased estimator of $\ell_i - \ell_{i'}$, $\delta_{\mathscr{P}} = Y_{i1} - \psi_{\mathscr{P}} - Y_{i'2}$, *with*

$$Var(\delta_{\mathscr{P}}) = \sigma_{i1}^2 + \sigma_{i'2}^2 + \sum_{k=0}^{M} (\sigma_{i_k j_k}^2 + \sigma_{i_k j_{k+1}}^2),$$

and for any other path $\mathscr{Q}$ different from $\mathscr{P}$,

$$Cov(\delta_{\mathscr{P}}, \delta_{\mathscr{Q}}) = \sigma_{i1}^2 + \sigma_{i'2}^2 + \sum_{k,r} \delta_{i_k i_r} [\delta_{j_k j_r} \sigma_{i_k j_k}^2 - \delta_{j_{k+1} j_r} \sigma_{i_k j_{k+1}}^2 - \delta_{j_k j_{r+1}} \sigma_{i_k j_k}^2 + \delta_{j_{k+1} j_{r+1}} \sigma_{i_k j_{k+1}}^2].$$

Under model (4) the term $\beta_i^2 + \beta_{i'}^2$ should be added to the variance and the covariance. Any convex combination, $\sum_{\mathscr{P}} w_{\mathscr{P}} \delta_{\mathscr{P}}$ (or $\sum_{\mathscr{P}} w_{\mathscr{P}} \psi_{\mathscr{P}}$), is also an unbiased estimator. The vector of estimated weights

$$\hat{w} = \frac{\widehat{\Sigma}^{-1} \mathbf{e}}{\mathbf{e}^{\mathrm{T}} \widehat{\Sigma}^{-1} \mathbf{e}},$$

where $\widehat{\Sigma}$ is an estimate of the covariance matrix Σ of the vector $\delta_{\mathscr{P}}$ formed by all paths $\mathscr{P}$, with the diagonal elements

$$\widehat{Var}(\delta_{\mathscr{P}}) = s_{i1}^2 + s_{i'2}^2 + \sum_{k=0}^{M} (s_{i_k j_k}^2 + s_{i_k j_{k+1}}^2),$$

and similarly defined off-diagonal elements, leads to a Graybill–Deal type estimator.

The proof of Proposition 9 is straightforward. Notice that when the covariance matrix Σ is known and is nonsingular, the estimator based on weights $\Sigma^{-1}\mathbf{e}/(\mathbf{e}^{\mathrm{T}}\Sigma^{-1}\mathbf{e})$ is UMVUE. This fact follows from the general Gauss–Markov Theorem.

One can assume in Proposition 9 that $\mathscr{P}$ is a minimal path, i.e. it does not contain a proper subset which is another path (as non-minimal paths have larger variances). In Example 4 there are four minimal paths:

$$\mathscr{P}_1 = \begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix}, \quad \mathscr{P}_2 = \begin{pmatrix} 1 & 4 \\ 1 & 3 \end{pmatrix}, \quad \mathscr{P}_3 = \begin{pmatrix} 2 & 3 \\ 1 & 3 \end{pmatrix}, \quad \mathscr{P}_4 = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}.$$

For the artifact contrast $a_1 - a_2$, any convex combination of the corresponding four (dependent) unbiased estimators

$$\psi = w_1 \psi_{\mathscr{P}_1} + w_2 \psi_{\mathscr{P}_2} + w_3 \psi_{\mathscr{P}_3} + w_4 \psi_{\mathscr{P}_4} = (w_1 + w_2)(Y_{11} - Y_{13}) + (w_3 + w_4)(Y_{21} - Y_{23}) + (w_1 + w_3)(Y_{33} - Y_{32}) + (w_2 + w_4)(Y_{43} - Y_{42})$$

has form (10) with $w = w_1 + w_2$ and $u = w_1 + w_3$. Thus, ψ can be conveniently represented as a linear combination of four independent statistics $Y_{ij} - Y_{ij'}$.

If, as in Examples 2 and 3, $G_1 \cap G_2 \neq \emptyset$, say $L_I \in G_1 \cap G_2$, then $Y_{I1} - Y_{I2}$ is always an unbiased estimator of $a_1 - a_2$. When $J = 2$, i.e. if there are only two artifacts, then any minimal path leads to such an unbiased estimator of $a_1 - a_2$.

6. Confidence intervals and simulation results

Cox (1975) gives a partially Bayesian setting of this problem in model (1) in which all unknown variances σ_{ij}^2 are independent and have common prior inverse-gamma density of form (8). Then conditionally on the sample variances s_{ij}^2, the estimator $\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j)$ has a distribution of the form

$$\sum_j \sum_{i \in G_j} b_{ij}(\ell_i + a_j) + \sum_j \sum_{i \in G_j} b_{ij} \left[\frac{n_{ij}\kappa_{ij}s_{ij}^2 + 0.5/\lambda}{n_{ij}(\kappa_{ij} + \kappa)} \right]^{1/2} t_{\kappa_{ij}+\kappa}.$$

Here $\kappa = 2\alpha$ and λ are parameters of the prior distribution and $t_{\kappa_{ij}+\kappa}$'s are independent t-distributed random variables with indicated degrees of freedom. The same result holds when the parameters of the prior inverse-gamma densities of σ_{ij}^2 are different with obvious modifications.

Since all such (conditional) distributions have the form $\sum c_k t_{v_k}$, they may be approximated by a multiple $\hat{c}t_{\hat{v}}$ of a t-random variable as follows:

$$\hat{v} = 4 + \frac{(\sum c_k^2(v_k/(v_k - 2)))^2}{\sum c_k^4(v_k^2/(v_k - 2)^2(v_k - 4))}, \tag{11}$$

$$\hat{c} = \sqrt{\frac{\hat{v} - 2}{\hat{v}} \sum c_k^2 \frac{v_k}{v_k - 2}}, \tag{12}$$

(Welch, 1937; Cox, 1975, p. 653; Fairweather, 1972, p. 231). This approximation is derived by equating the second and the fourth cumulants and can be used only if $\min v_k > 4$. If $2 < \min v_k \leqslant 4$, an analysis of the characteristic function of t-distribution suggests the following formula for the degrees of freedom:

$$\bar{v} = 2 + \frac{\left(\sum c_k \sqrt{v_k}\right)^2}{\sum c_k^2 \frac{v_k}{v_k - 2}},$$

and for the multiple c,

$$\bar{c} = \frac{\sum c_k \sqrt{v_k}}{\sqrt{\bar{v}}}.$$

When there is some prior information, which allows specification of the parameters α and λ (for example, as discussed in Section 4), one can use this approximation to obtain a confidence interval as indicated below. The "non-informative" prior corresponds to $\kappa = 0$ and $\lambda = \infty$. In other terms, the generalized prior for σ_{ij} is $d\sigma_{ij}/\sigma_{ij}$. In this case the above development leads to the same results as the fiducial distribution (Lee, 1997), i.e. the distribution can be derived from Fisher's fiducial argument.

As relevant prior information in the existing linkage studies seems to be missing, we used the non-informative prior in all our examples for artifact contrasts. In Example 1, say, with $I = 1$, the conditional distribution of $(Y_{11} - Y_{12} - a_1 + a_2)/\sqrt{s_{11}^2 + s_{12}^2}$ for the fixed ratio s_{11}^2/s_{12}^2 is the linear combination with the weights $\omega_1 = s_{11}/\sqrt{s_{11}^2 + s_{12}^2}$ and $\sqrt{1 - \omega_1^2} = s_{12}/\sqrt{s_{11}^2 + s_{12}^2}$ of two independent t-distributions with degrees of freedom κ_{11} and κ_{12}, respectively, i.e. this is exactly the classical Behrens–Fisher distribution (Lee, 1997, pp. 148, 289–300).

Patil (1965) has studied the accuracy of the approximation through (11) and (12) and has found it superior to the Welch–Satterthwaite approximation discussed in Section 3 (which takes $c = 1$ and v from (3)). See also Fenstad (1983) and Best and Rayner (1987). Fig. 1 shows that these two approximations are really very close. For this reason (and

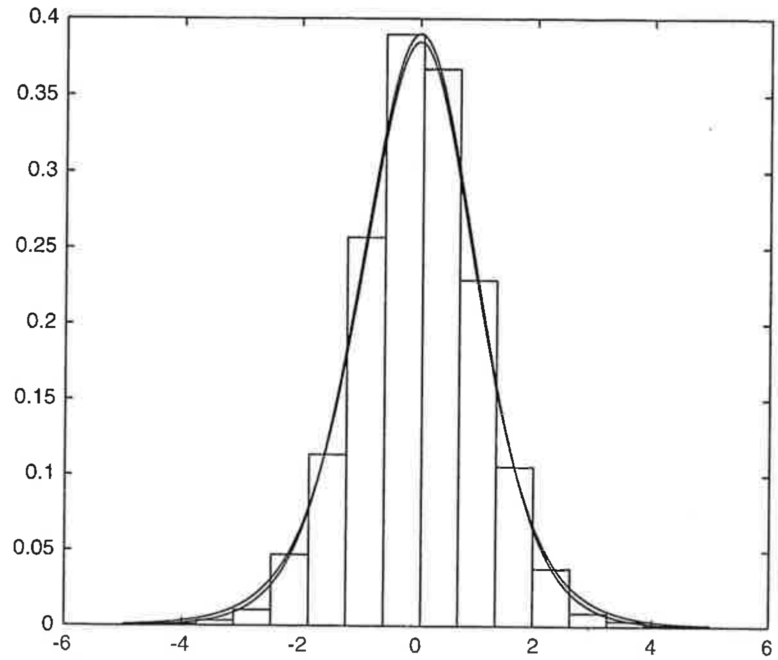

Fig. 1. The histogram of the distribution of $(Y_{11} - Y_{12} - a_1 + a_2)/\sqrt{s_{11}^2 + s_{12}^2}$ and its t-density approximations (via (11) and (12), and the Welch–Satterthwaite formula) in Example 1.1.

since the Welch–Satterthwaite approximation is not applicable in the incomplete case) we report later numerical results only for the more versatile multiple of t-variable approximation given in formulas (11) and (12).

To derive an (approximate) confidence intervals for $a_1 - a_2$ in Example 2, observe that the conditional distribution of the pivotal ratio, $(\psi - a_1 + a_2)/\sqrt{\widehat{Var(\psi)}}$ for fixed $\hat{w}$, ω_1 and $\omega_2 = s_{21}/\sqrt{s_{21}^2 + s_{22}^2}$, has the form $\sqrt{\hat{w}}(\omega_1 t_{\kappa_{11}} + \sqrt{1-\omega_1^2} t_{\kappa_{12}}) + \sqrt{1-\hat{w}}(\omega_2 t_{\kappa_{21}} + \sqrt{1-\omega_2^2} t_{\kappa_{22}})$. In other terms, this conditional distribution is a mixture of four independent t-distributions, which can be approximated by a multiple of another t-distribution via (11) and (12). A similar representation (with six independent t-distributions) holds in the Example 3 and in Example 4 for statistic (10) with estimated weights $\hat{w}$ and $\hat{u}$.

There are numerous studies of estimators of the common mean (see Rao, 1980; Rao et al., 1981; Rukhin and Vangel, 1998.) Approximate confidence intervals for the common mean were suggested in Rukhin (2003). In our situation the role of measurements to be combined is played by the differences $Y_{I1} - Y_{I2}$ with the variances $\sigma_{I1}^2 + \sigma_{I2}^2$, or even more complicated (dependent) sums of differences $\sum_k (Y_{i_k j_k} - Y_{i_k j_{k+1}})$. Still, the desired confidence intervals can be obtained from the conditional distribution of $(\psi_w - a_1 + a_2)/\sqrt{\widehat{Var(\psi_w)}}$ via its t-approximation after (11) and (12).

Monte Carlo simulation results on the coverage probability of the confidence intervals for the between-artifacts contrast $a_1 - a_2$ in Examples 1–4 are given in Fig. 2. We have taken the sample sizes of K pilot labs to be random permutations of integers 1 through K plus four. The values of K and n are: $K = 1, n = 2$ in Example 1.1, $K = 2, n = 4$ in Example 1.3, $K = 2, n = 4$ in Example 2, $K = 3, n = 6$ in Example 3, and $K = 4, n = 8$ in Example 4. The variances σ_{ij}^2 were taken to be a realization of an inverse gamma-variable (8) with $\alpha = 2$ for the scale parameter λ taking values $0.1 : 0.5 : 3$.

Clearly, all five intervals (especially in Example 1.1) have a deteriorating coverage probability for larger values of λ, but can be safely used for smaller values of this parameter, say, when $\lambda < 1.5$. The standard errors (the average half-widths) of these intervals are increasing as λ increases but not fast enough to attain the desired coverage, which confirms the deteriorating performance of the confidence intervals for large error variance.

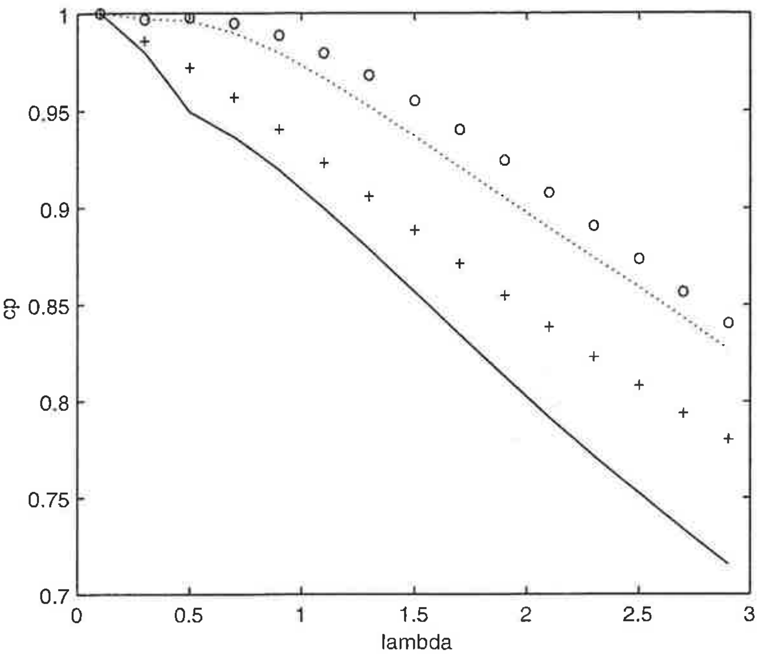

Fig. 2. Plot of coverage probability of confidence intervals vs λ for the Examples 1–4 (the continuous line corresponds to Example 1.1, "+" line to Example 1.3, "*" line to Example 2, the dotted line ":" to Example 3, the line marked by "o" to Example 4).

7. Conclusions

This paper is concerned with the issue of linkage in interlaboratory studies. We have shown that when each pilot lab measures exactly two artifacts, the usual sufficient statistic is complete only when the number of artifacts is one greater than the number of pilot labs. It is demonstrated that UMVUE of certain contrasts exist even in the non-saturated case. Additionally, we have modeled and investigated the effect of Type B error on the above consideration and have produced Welch–Satterthwaite and partially Bayesian confidence intervals for the contrasts.

In most practical situations, one has to link two KC studies. Thus, with $J=2$, an estimator of a_1-a_2 is desired. For this purpose all pilot labs participating in both studies are to be determined. Each such lab, $L_I \in G_1 \cap G_2$, leads to an unbiased estimator $Y_{I1}-Y_{I2}$ of a_1-a_2. All these independent estimators can be combined to obtain the Graybill–Deal estimator $\psi=\sum \hat{w}_k(Y_{I_k1}-Y_{I_k2})$ with $\hat{w}_k=(s^2_{I_k1}+s^2_{I_k2})^{-1}/(\sum_m(s^2_{I_m1}+s^2_{I_m2})^{-1})$. An approximate t-confidence interval based on the non-informative prior for the variances can be obtained from the conditional distribution of the pivotal ratio, $(\psi-a_1+a_2)/\sqrt{\widehat{Var(\psi)}}$ for fixed $\hat{w}_k$, $\omega_k=s_{I_k1}/\sqrt{s^2_{I_k1}+s^2_{I_k2}}$, has the form $\sum_k \sqrt{\hat{w}_k}(\omega_k t_{\kappa_{I_k1}}+\sqrt{1-\omega_k^2}\,t_{\kappa_{I_k2}})$. This distribution can be approximated by a multiple of a t-distribution via (11) and (12).

The situation is more complicated when $J>2$ and the linkage must be made through multiple pilot labs or when there are no pilot labs participating in both studies, i.e. when the set $G_1 \cap G_2$ is empty. Then all (minimal) paths connecting the labs measuring the artifacts of interest are to be enumerated and combined as in Proposition 9. An unbiased estimator for degrees of equivalence, $\ell_i-\ell_{i'}$, is always obtainable as $Y_{i1}-\psi_m-Y_{i'2}$. This interval involves Type B errors $\beta_i^2+\beta_{i'}^2$ which must be added to the variance term under model (4).

Acknowledgements

This work has been done at the Statistical Engineering Division, National Institute of Standards and Technology. The authors are grateful to Nell Sedransk and Will Guthrie for stimulating discussions.

Appendix. Proofs

Proof of Proposition 2. In the context of Proposition 1, the dimension $M+K$ in our situation is m. Indeed, $M-K$ non-pilot laboratories perform exactly one measurement, and K pilot laboratories each perform two measurements. A standard argument of analysis of variance shows that the maximum possible dimension of the vector space Θ is $M+J-1$. Hence, according to Proposition 1, incompleteness holds when $J-1<K$.

Each non-pilot laboratory contributes 1 to the dimension of the space spanned by the vector $\ell_i + a_j$, $i \in G_j$, $j=1,\ldots,J$, since ℓ_i is a free parameter unique for this lab. Also the equivalence relation is not affected by addition or removal of a non-pilot lab. Hence it suffices to prove Proposition 2 for an experiment consisting only of K pilot laboratories $L_1,\ldots,L_K$. There must exist a maximal path, which one can take to be $L_1 \sim L_2 \sim \cdots \sim L_K$. If a_1 and a_2 are two artifacts measured by L_1, then at least one of them, say, a_2 must be measured by L_2 as these two labs are directly linked. Since each lab in the linked sequence measures at least one artifact processed by the previous lab, each succeeding lab must measure one artifact not already dealt with by any previous lab in the sequence. Otherwise, the number of different measured artifacts would be less than $K+1$, i.e. each lab starting with L_2 can process at most one artifact not already measured, and it must measure at least one new artifact if the total number of measured artifacts is $K+1$. (This argument also shows that the range of possible values of J is from 2 to $K+1$.)

Thus, the artifacts can be ordered as $a_1,\ldots,a_{K+1}$, where a_{j+1} is the new artifact measured by L_j, but not processed by any previous lab in the sequence. This leads to the ordering of $2K$ means $\mu_1,\ldots,\mu_{2K}$, as $\mu_1=\ell_1+a_1$, $\mu_2=\ell_1+a_2,\ldots,\mu_{2i-1}=\ell_i+a$, where $a\in\{a_1,\ldots,a_i\}$, and $\mu_{2i}=\ell_i+a_{i+1}$, $i=2,\ldots,K$. Hence the dimension of the space spanned by $(\mu_1,\ldots,\mu_{2K})$ is $2K$ since each successive coordinate includes a free parameter. □

Proof of Proposition 5. Notice first that the data (Y,S^2) is sufficient for itself. It follows from independence of all variables and the definition of sufficiency that the vector (X,S^2) is sufficient for "augmented" data (B,X,S^2), and, by Proposition 1, is complete when $p=m$. In this case assume that for some function $h(Y,S^2)$, $Eh(Y,S^2)=0$ for all parametric values. Then by completeness, of (X,S^2), $E(h(B+X,S^2)|X,S^2)=0$ almost surely. Hence for almost every fixed value of (x,s^2), $Eh(B+x,s^2)=0$. However, the distribution of $B+x$ for $x\in R^m$ is always complete, so that $h(x,s^2)=0$ with probability one, i.e. (Y,S^2) is a complete statistic.

Incompleteness of (Y,S^2) when $k<m$ follows from Proposition 1 since the expected values of linear functions of Y and X are identical. □

Proof of Proposition 7. Under our assumptions X is sufficient for (X,Y) when $\eta=\eta^*$. Suppose $\delta(X,Y)$ is any unbiased estimator of $g(\theta)$. Then by the Rao–Blackwell Theorem and the fact that $\delta(X)$ is UMVUE of $g(\theta)$, when only X is observed, it follows that $Var_{\theta,\eta^*}\delta(X,Y)\geqslant Var_{\theta,\eta^*}E_{\eta^*}(\delta(X,Y)|X)\geqslant Var_{\theta,\eta^*}(\delta(X))$. Since η^* was arbitrary, this inequality holds for all (θ,η), and hence the result follows. □

Proof of Proposition 8. Partition the data into $(X_1,\ldots,X_r,S_1^2,\ldots,S_r^2)$, $(X_{r+1},\ldots,X_m,S_{r+1}^2,\ldots,S_m^2)$. The first set is sufficient and complete for $(\mu_1,\ldots,\mu_r,\sigma_1^2,\ldots,\sigma_r^2)$ in the absence of $(X_{r+1},\ldots,X_m,S_{r+1}^2,\ldots,S_m^2)$. Hence $\delta(X_1,\ldots,X_r,S_1^2,\ldots,S_r^2)$ is UMVUE of its expected value if just $(X_1,\ldots,X_r,S_1^2,\ldots,S_r^2)$, is observed. By Proposition 7, $\delta(X_1,\ldots,X_r,S_1^2,\ldots,S_r^2)$ is UMVUE in the model with full data. □

References

Best, G.J., Rayner, J.C.W., 1987. Welch's approximate solution to the Behrens–Fisher problem. Technometrics 29, 205–210.

Bureau International des Poids et Mesures, 1999. Mutual Recognition of National Measurement Standards and Measurement Certificates Issued by National Metrology Institutes. Technical Report, Sevres, France.

Cox, D.R., 1975. A note on partially Bayes inference and the linear model. Biometrika 62, 651–654.

Delahaye, F., Witt, T.J., 2002. Linking the results of Key Comparison CCEM-K4 with the 10 pF results of EUROMET Project 345. Metrologia, 39, 01005 (Technical Supplement).

Elster, C., Link, A., Woeger, W., 2003. Proposal for linking the results of CIPM and RMO key comparisons. Metrologia 40, 189–194.

Fairweather, W.R., 1972. A method for obtaining an exact confidence interval for the common mean of several normal populations. Appl. Statist. 21, 229–233.

Fenstad, G.U., 1983. A comparison between the U and V tests in the Behrens–Fisher problem. Biometrika 70, 300–302.

Graybill, F.A., Deal, R.D., 1959. Combining unbiased estimators. Biometrics 15, 543–550.

International Organization for Standardization ISO, 1993. Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization ISO, Geneva, Switzerland (corrected and reprinted 1995).

Lee, P., 1997. Bayesian Statistics: An Introduction. Arnold, London.

Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.

Linnik, Yu.V., 1968. Statistical Problems with Nuisance Parameters. American Mathematical Society, Providence RI.

Magnum, B.W., Strouse, G.F., Guthrie, W.F., 2002. CCT-K3. Key Comparison of Realizations of the ITS-90 over the Range 83.8058K to 933.473K. NIST Technical Note 1450, Gaithersburg.

Patil, V.H., 1965. Approximation to the Behrens–Fisher distribution. Biometrika 52, 267–271.

Rao, J.N.K., 1980. Estimating the common mean of possibly different normal populations: a simulation study. J. Amer. Statist. Assoc. 75, 447–453.

Rao, P.S.R.S., Kaplan, J., Cochran, W.G., 1981. Estimators for the one-way random effects model with unequal error variances. J. Amer. Statist. Assoc. 76, 89–97.

Rukhin, A.L., 2003. Two procedures of meta-analysis in clinical trials and interlaboratory studies. Tatra Mountains Math. Publ. 28, 155–168.

Rukhin, A.L., Vangel, M.G., 1998. Estimation of a common mean and weighted means statistics. J. Amer. Statist. Assoc. 93, 303–308.

Taylor, B.N., Kuyatt, C.E., 1994. Guidelines for Evaluating and Expressing Uncertainty in NIST Measurement Results, NIST TN 1297, NIST, Gaithersburg, MD.

Voinov, V.G., Nikulin, M.S., 1993. Unbiased Estimators and Their Applications. Kluwer Academic Publishers, Dordrecht.

Welch, B.L., 1937. The significance of the difference between the means when the population variances are unequal. Biometrika 29, 350–362.