INSTITUTE OF PHYSICS PUBLISHING METROLOGIA
Metrologia **41** (2004) 132–136 PII: S0026-1394(04)73994-8

# Combining information from interlaboratory evaluations using a random effects model

**Raghu N Kacker**

National Institute of Standards and Technology, Gaithersburg, MD 20899-8910, USA

E-mail: raghu.kacker@nist.gov

Received 22 October 2003
Published 20 February 2004
Online at stacks.iop.org/Met/41/132 (DOI: 10.1088/0026-1394/41/3/004)

**Abstract**
This paper compares leading methods for combining information from interlaboratory evaluations of a common measurand through a random effects model of classical statistics. The leading methods are those of Cochran, Paule and Mandel, and DerSimonian and Laird. We show that all three methods are special cases of a unifying identity. The unifying identity suggests a new two-step method. This makes four methods for comparison. The comparison is based on six published data sets from three key comparisons. The method of Paule and Mandel is optimal in the sense of being conditionally restricted maximum likelihood under normality, the condition being that the estimated intralaboratory variances be treated as the true variances. The method of Paule and Mandel requires a simple iteration that can be easily done on a spreadsheet program. Therefore, it is the preferred method for combining results of interlaboratory evaluations through a random effects model. We compare the other three methods relative to the method of Paule and Mandel. The two-step method approximates the optimal method of Paule and Mandel better than the earlier methods of Cochran, and DerSimonian and Laird.

## 1. Introduction

A generic problem in combining information from interlaboratory evaluations is as follows. We are given some number $m$ of individual laboratory results $x_1, \ldots, x_m$ and their associated standard uncertainties, $s(x_1), \ldots, s(x_m)$, in measurement of the value $\mu$ of a common measurand. The results $x_1, \ldots, x_m$ are often arithmetic means that have been corrected (adjusted) for recognized systematic effects in the individual laboratories. The uncertainties $s(x_1), \ldots, s(x_m)$ include components of uncertainty associated with the corrections. The objective of combining information is to determine a combined result $x_C$ for $\mu$ and its associated standard uncertainty, $s(x_C)$, based on the data $x_1, \ldots, x_m$ and $s(x_1), \ldots, s(x_m)$.

An assumed model[1] for the relationship between the data and $\mu$ is required for determining $x_C$ and $s(x_C)$. A classical statistics approach is to assume a random effects model [1]. The results $x_1, \ldots, x_m$ are assumed to be realizations of random variables[2] having normal distributions with expected values $\mu_1, \ldots, \mu_m$ and variances $\sigma_1^2, \ldots, \sigma_m^2$, respectively, and $s^2(x_1), \ldots, s^2(x_m)$ are regarded as estimates of $\sigma_1^2, \ldots, \sigma_m^2$, respectively. The variable $x_i$ may be parsed as $x_i = \mu_i + e_i$, where $e_i$ is the normally distributed random error $(x_i - \mu_i)$ with expected value zero and variance $\sigma_i^2$ for $i = 1, 2, \ldots, m$. The errors $e_1, \ldots, e_m$ are assumed to be independently distributed. We refer to $\mu_1, \ldots, \mu_m$ as laboratory expected values and refer to $\sigma_1^2, \ldots, \sigma_m^2$ as intralaboratory variances. The variables $x_1, \ldots, x_m$ are related to the value $\mu$ of the common measurand by the following model:

$$x_i = \mu_i + e_i = \mu + (\mu_i - \mu) + e_i = \mu + b_i + e_i, \qquad (1)$$

where $b_i$ is the bias $(\mu_i - \mu)$ in $x_i$ for $i = 1, 2, \ldots, m$. The model (1) constructed so far does not specify the relationship

[1] The statistical conclusions are conditional on the assumed model. Therefore, the conclusions are justified only to the extent that the assumed model is justified.

[2] The symbols $x_1, \ldots, x_m$ are used for both the results and the corresponding random variables.

  

between the data and $\mu$. The classical statistics assumptions[3] to relate the data to $\mu$ are as follows: the laboratory biases, $b_1, \ldots, b_m$, are assumed to be random variables having the same normal distribution with expected value zero and variance $\sigma_b^2 \geqslant 0$, called interlaboratory variance; and $b_1, \ldots, b_m$ are assumed to be mutually independent and independent of the errors, $e_1, \ldots, e_m$. The numbers $n_1, \ldots, n_m$ of observations used to determine the laboratory results $x_1, \ldots, x_m$ may not be equal. So the model (1) may be unbalanced. The intralaboratory variances, $\sigma_1^2, \ldots, \sigma_m^2$, may be different. So the model (1) may be heteroscedastic. There could be many reasons for different variances $\sigma_1^2, \ldots, \sigma_m^2$ including unequal $n_1, \ldots, n_m$. According to the random effects model (1), the expected value, $E(x_i)$, is $\mu$, the variance, $V(x_i)$, is $\sigma_b^2 + \sigma_i^2$, and $x_i$ is normally distributed for $i = 1, 2, \ldots, m$. The model (1) accounts for the laboratory biases $b_1, \ldots, b_m$ by introducing a new parameter, $\sigma_b^2$, the interlaboratory variance component of the variances $V(x_1), \ldots, V(x_m)$.

In determining $x_C$ and $s(x_C)$, it is difficult to quantify and incorporate the uncertainty that arises from the use of estimates $s^2(x_1), \ldots, s^2(x_m)$ for $\sigma_1^2, \ldots, \sigma_m^2$. Therefore, in conjunction with model (1), many metrologists make the following simplifying assumption: the estimates $s^2(x_1), \ldots, s^2(x_m)$ are regarded as the true variances $\sigma_1^2, \ldots, \sigma_m^2$. The conditional variances of $x_1, \ldots, x_m$ under this assumption are $\sigma_b^2 + s^2(x_1), \ldots, \sigma_b^2 + s^2(x_m)$, respectively. The conditional standard deviation of the combined result, $x_C$, based on this assumption is necessarily an underestimate of its unconditional standard deviation [2]. This paper is based on the simplifying assumption. Thus all statistical analyses and properties discussed here are conditional.

Many metrologists use the following weighted mean, $x_W$, as the combined result, $x_C$:

$$x_W = \frac{\sum_i w_i x_i}{\sum_i w_i}, \tag{2}$$

where $w_i = 1/(s_b^2 + s^2(x_i))$, for $i = 1, 2, \ldots, m$, and $s_b^2$ is an estimate of $\sigma_b^2$ determined from the available data. The maximum-likelihood (ML) estimate of the parameter $\mu$ of model (1), conditional on the simplifying assumption, is $x_W(\sigma_b) = \sum_i W_i x_i / \sum_i W_i$, where $W_i = 1/(\sigma_b^2 + s^2(x_i))$. Thus $x_W$ of equation (2) is a good estimate of $\mu$ when the random effects model (1) and the other assumptions are justified and $s_b^2$ is a good estimate of $\sigma_b^2$. Many metrologists use the quantity $1/\sqrt{\sum_i w_i}$, where $w_i = 1/(s_b^2 + s^2(x_i))$, as the estimated standard deviation, $s(x_W)$, of $x_W$ (Paule and Mandel [3] and DerSimonian and Laird [4]). The quantity $1/\sqrt{\sum_i w_i}$ is an underestimate of the conditional standard deviation of $x_W$ because it does not include the component of uncertainty that arises from the use of estimate $s_b^2$ for $\sigma_b^2$.

The objective of this paper is to compare leading methods to determine an estimate, $s_b^2$, for $\sigma_b^2$ and the corresponding combined result, $x_W$, of equation (2). The leading methods are those of Cochran [5], Paule and Mandel [3], and DerSimonian and Laird [4]. Cochran's method is based on analysis-of-variance (ANOVA) of the data. Paule and Mandel's method and its modifications have often been used to certify Standard Reference Materials at the National Institute of Standards and Technology (NIST) (Schiller and Eberhardt [6]). DerSimonian and Laird's method is a popular method for combining information from clinical trials, where the clinical trials take the role of interlaboratory evaluations[4]. Paule and Mandel's method requires a simple iteration. DerSimonian and Laird's method seems to be popular, in part because it is non-iterative. We show that all three methods are special cases of a unifying identity. The unifying identity suggests a new two-step method. Together with the two-step method, we have a pool of four methods for comparison. We use six data sets from three key comparisons to compare the four methods for estimating $\sigma_b^2$. Rukhin *et al* [7] show that the estimate of Paule and Mandel is optimal in the sense that, under the simplifying assumption, it is a restricted ML (REML) estimate of $\sigma_b^2$. So we compare the other three methods relative to the method of Paule and Mandel.

## 2. Unifying identity

Cochran's ANOVA estimate for $\sigma_b^2$ is

$$s_b^2(\mathrm{CA}) = \max\left\{0, \frac{1}{m-1}\sum_i (x_i - x_A)^2 - \frac{1}{m}\sum_i s^2(x_i)\right\}, \tag{3}$$

where $x_A = (1/m)\sum_i x_i$ is the arithmetic mean of the results. The corresponding combined result, $x_W(\mathrm{CA})$, is obtained by substituting $s_b^2(\mathrm{CA})$ for $s_b^2$ in equation (2).

The estimate proposed by Paule and Mandel is the solution $\sigma_b^2$ of the following estimating equation:

$$F(\sigma_b^2) = \sum_i W_i (x_i - x_W(\sigma_b))^2 - (m-1) = 0, \tag{4}$$

where $W_i = 1/(\sigma_b^2 + s^2(x_i))$, for $i = 1, \ldots, m$, and $x_W(\sigma_b) = \sum_i W_i x_i / \sum_i W_i$. The solution, $s_b^2(\mathrm{PM})$, of equation (4) is determined through a simple iteration. When $F(\sigma_b^2) < 0$ for all $\sigma_b^2 \geqslant 0$, the estimate, $s_b^2(\mathrm{PM})$, is set to zero. The corresponding combined result, $x_W(\mathrm{PM})$, is obtained by substituting $s_b^2(\mathrm{PM})$ for $s_b^2$ in equation (2).

The estimate proposed by DerSimonian and Laird for $\sigma_b^2$ is

$$s_b^2(\mathrm{DL}) = \max\left\{0, \frac{[\sum_i w_{i0}(x_i - x_w(0))^2] - (m-1)}{\sum_i w_{i0} - \sum_i w_{i0}^2 / \sum_i w_{i0}}\right\}, \tag{5}$$

where $w_{i0} = 1/s^2(x_i)$, for $i = 1, 2, \ldots, m$, and $x_W(0) = \sum_i w_{i0} x_i / \sum_i w_{i0}$. That is, $w_{i0}$ and $x_W(0)$ are obtained by substituting 0 for $s_b^2$ in the equation $w_i = 1/(s_b^2 + s^2(x_i))$ and equation (2), respectively. The corresponding combined result, $x_W(\mathrm{DL})$, is obtained by substituting $s_b^2(\mathrm{DL})$ for $s_b^2$ in equation (2).

A two-step estimate for $\sigma_b^2$ is as follows. First, compute $s_b^2(\mathrm{CA})$ and $x_W(\mathrm{CA})$. Then compute the following

[3] These are very strong assumptions that may or may not be reasonable for the particular data.

[4] As of 15 October 2003, Paule and Mandel's 1982 paper has been cited 27 times and DerSimonian and Laird's 1986 paper has been cited 1543 times.

estimate for $\sigma_b^2$,

$$s_b^2(\mathrm{C2}) = \max\left\{0, \left(\left[\sum_i w_{iC}(x_i - x_w(\mathrm{CA}))^2\right] - \left[\sum_i w_{iC}s^2(x_i) - \frac{\sum_i w_{iC}^2 s^2(x_i)}{\sum_i w_{iC}}\right]\right) \times \left[\sum_i w_{iC} - \frac{\sum_i w_{iC}^2}{\sum_i w_{iC}}\right]^{-1}\right\}, \quad (6)$$

where $w_{iC} = 1/(s_b^2(\mathrm{CA}) + s^2(x_i))$ for $i = 1, 2, \ldots, m$. The corresponding combined result, $x_w(\mathrm{C2})$, is obtained by substituting $s_b^2(\mathrm{C2})$ for $s_b^2$ in equation (2). The symbol C2 in $s_b^2(\mathrm{C2})$ indicates that it is a two-step estimate based on Cochran's ANOVA estimate, $s_b^2(\mathrm{CA})$.

It turns out that all four methods are special cases of the following unifying identity. If $x_1, \ldots, x_m$ are independent random variables with the same expected value, $\mu$, and variances $\sigma_b^2 + \sigma_1^2, \ldots, \sigma_b^2 + \sigma_m^2$, respectively, $\sigma_1^2 > 0, \ldots, \sigma_m^2 > 0$, $a_1, \ldots, a_m$ are positive constants, and $x_C = \sum_i a_i x_i / \sum_i a_i$, then

$$E\left[\sum_i a_i(x_i - x_C)^2\right] = \sum_i a_i(\sigma_b^2 + \sigma_i^2) - \frac{\sum_i a_i^2(\sigma_b^2 + \sigma_i^2)}{\sum_i a_i}. \quad (7)$$

Alternatively,

$$E\left[\sum_i a_i(x_i - x_C)^2\right] = \sigma_b^2\left[\sum_i a_i - \frac{\sum_i a_i^2}{\sum_i a_i}\right] + \left[\sum_i a_i\sigma_i^2 - \frac{\sum_i a_i^2\sigma_i^2}{\sum_i a_i}\right]. \quad (8)$$

This identity can be easily verified (appendix A). By substituting $s^2(x_1), \ldots, s^2(x_m)$ for $\sigma_1^2, \ldots, \sigma_m^2$ in equation (8), we get the following method-of-moments (MM) estimate, $s_b^2(\mathrm{MM})$, for $\sigma_b^2$:

$$s_b^2(\mathrm{MM}) = \left(\left[\sum_i a_i(x_i - x_C)^2\right] - \left[\sum_i a_i s^2(x_i) - \frac{\sum_i a_i^2 s^2(x_i)}{\sum_i a_i}\right]\right)\left[\sum_i a_i - \frac{\sum_i a_i^2}{\sum_i a_i}\right]^{-1}. \quad (9)$$

Since $\sigma_b^2 \geqslant 0$, the estimate $s_b^2(\mathrm{MM})$ is legitimate only when it is non-negative. When equation (9) is negative, we set $s_b^2(\mathrm{MM}) = 0$.

*Estimate* $s_b^2(\mathrm{CA})$: If we substitute $a_i = 1/m$ for $i = 1, \ldots, m$ in equation (9), we get Cochran's estimate, $s_b^2(\mathrm{CA})$.

*Estimate* $s_b^2(\mathrm{DL})$: If we substitute $a_i = 1/s^2(x_i)$ for $i = 1, \ldots, m$ in equation (9), we get DerSimonian and Laird's estimate, $s_b^2(\mathrm{DL})$.

*Estimate* $s_b^2(\mathrm{C2})$: If we substitute $a_i = 1/(s_b^2(\mathrm{CA}) + s^2(x_i))$ for $i = 1, \ldots, m$ in equation (9), we get the two-step estimate, $s_b^2(\mathrm{C2})$.

*Estimate* $s_b^2(\mathrm{PM})$: If we substitute $a_i = 1/(\sigma_b^2 + \sigma_i^2)$ for $i = 1, 2, \ldots, m$ in equation (7), we get $E[\sum_i a_i(x_i - x_C)^2] = m - 1$. Paule and Mandel's estimating equation (4) is then obtained by equating $\sum_i a_i(x_i - x_C)^2$ to its expected value, $m - 1$, where $a_i = 1/(\sigma_b^2 + \sigma_i^2)$, and then substituting $s^2(x_1), \ldots, s^2(x_m)$ for $\sigma_1^2, \ldots, \sigma_m^2$.

## 3. Method of Paule and Mandel and its optimality

The method of Paule and Mandel to determine $s_b^2(\mathrm{PM})$ from the estimating equation (4) is the classical Newton's method of calculus for approximating the zeros of real-valued functions. The algorithm is as follows. Start with $\sigma_b^2(\mathrm{previous}) = 0$ or with a number slightly above zero.

(i) Calculate weights $W_i = 1/(\sigma_b^2 + s^2(x_i))$ for $i = 1, 2, \ldots, m$ and the function $F(\sigma_b^2)$.

(ii) If $F(\sigma_b^2)$ at $\sigma_b^2 = 0$ is negative, set $s_b^2(\mathrm{PM}) = 0$.
If $F(\sigma_b^2(\mathrm{previous})) = 0$, set $s_b^2(\mathrm{PM}) = \sigma_b^2$ (previous).
If $F(\sigma_b^2(\mathrm{previous})) > 0$, determine the correction

$$\Delta\sigma_b^2 = \frac{\sum_i W_i(x_i - x_w(\sigma_b))^2 - (m-1)}{\sum_i W_i^2(x_i - x_w(\sigma_b))^2}. \quad (10)$$

(iii) The next iterative value of $\sigma_b^2$ is $\sigma_b^2(\mathrm{next}) = \sigma_b^2(\mathrm{previous}) + \Delta\sigma_b^2$.

(iv) Repeat (ii) and (iii) until $F(\sigma_b^2(\mathrm{previous})) = 0$. The final value of $\sigma_b^2$ is $s_b^2(\mathrm{PM})$.

This algorithm is simple enough to do on a spreadsheet program. Paule and Mandel suggest that the starting value of $\sigma_b^2$ should be set slightly above zero. We suggest $s_b^2(\mathrm{CA})$ as the starting value for $\sigma_b^2$. This often reduces the number of iterations required unless $s_b^2(\mathrm{CA})$ is zero. The algorithm of Paule and Mandel gives a unique solution for $s_b^2(\mathrm{PM})$ (appendix B).

Paule and Mandel did not assume that the errors $e_1, \ldots, e_m$ and biases $b_1, \ldots, b_m$ are normally distributed, and they did not investigate the statistical properties of $s_b^2(\mathrm{PM})$. Recently, Rukhin and Vangel [8] and Rukhin *et al* [7] investigated the properties of $s_b^2(\mathrm{PM})$ under normality. In particular, Rukhin *et al* show that when the errors $e_1, \ldots, e_m$ and biases $b_1, \ldots, b_m$ are normally distributed and a weighted mean of the form $x_w = \sum_i w_i x_i / \sum_i w_i$, where $w_i = 1/(s_b^2 + s^2(x_i))$, is used as an estimate for the value $\mu$ of the common measurand, the Paule and Mandel estimate $s_b^2(\mathrm{PM})$ is the conditionally REML estimate of $\sigma_b^2$, the condition being that the estimates $s^2(x_1), \ldots, s^2(x_m)$ be regarded as the true variances $\sigma_1^2, \ldots, \sigma_m^2$, respectively. A REML estimate is an improvement over the ML estimate of a variance component because it accounts for the loss in degrees of freedom resulting from estimation of $\mu$ [9]. Rukhin *et al* also show that the combined result, $x_w(\mathrm{PM})$, is an approximate generalized Bayes estimate based on non-informative prior distributions for the parameters $\mu, \sigma_1, \ldots, \sigma_m$, and $\sigma_b$. Thus, $s_b^2(\mathrm{PM})$ is an optimal estimate of the parameter $\sigma_b^2$ of model (1) under normality.

## 4. Comparison based on key comparison data

Rukhin [10] showed that the methods of DerSimonian and Laird, and Paule and Mandel are asymptotically similar. Rukhin's comparison does not apply when the number, $m$, of laboratories is less than 30, which is frequently the case. In order to compare the four methods for estimating $\sigma_b^2$, we have used six data sets, two from each of the three key comparisons labelled K2, K5, and K6 conducted by the International Consultative Committee on Amount of Substance (CCQM) (www.bipm.org). These data are suitable for comparison

**Table 1.** Estimates $s_b$(PM), $s_b$(CA), $s_b$(DL), and $s_b$(C2).

| | $s_b$(PM) | $s_b$(CA) | $s_b$(DL) | $s_b$(C2) |
|---|---|---|---|---|
| K2(Pb) | 0.8399 | 1.1837 | 0.5359 | 0.9352 |
| K2(Cd) | 0.3095 | 0.0000 | 0.4675 | 0.4675 |
| K5(N) | 0.0376 | 0.0365 | 0.0438 | 0.0377 |
| K5(F) | 0.1579 | 0.1530 | 0.1980 | 0.1582 |
| K6(A) | 0.0336 | 0.0339 | 0.0292 | 0.0336 |
| K6(B) | 0.0175 | 0.0206 | 0.0103 | 0.0181 |

**Table 2.** Estimates $x_W$(PM), $x_W$(CA), $x_W$(DL), and $x_W$(C2).

| | $x_W$(PM) | $x_W$(CA) | $x_W$(DL) | $x_W$(C2) |
|---|---|---|---|---|
| K2(Pb) | 62.4078 | 62.4438 | 62.3906 | 62.4175 |
| K2(Cd) | 82.9000 | 82.5357 | 83.0390 | 83.0390 |
| K5(N) | 1.5212 | 1.5111 | 1.5210 | 1.5212 |
| K5(F) | 5.9960 | 5.9960 | 5.9959 | 5.9960 |
| K6(A) | 2.1976 | 2.1976 | 2.1974 | 2.1976 |
| K6(B) | 1.7306 | 1.7310 | 1.7294 | 1.7307 |

because the results $x_1, \ldots, x_m$ are direct measurements of a common measurand, an amount of substance. The numbers of participating laboratories in these key comparisons are nine, ten, and seven, respectively. The comparison shown here indicates the differences one might expect when the number of laboratories is close to ten.

The measurands in the two data sets of the CCQM key comparison K2 are the amount of lead (Pb) and the amount of cadmium (Cd) in natural water measured in nmol kg$^{-1}$. The measurands in the two data sets of the CCQM key comparison K5 are the mass fraction of pp′-dichlorodiphenyl-dichloroethylene (DDE) in natural (N) and fortified (F) fish oils measured in µg g$^{-1}$. The measurands in the two data sets of the CCQM key comparison K6 are the mass fraction of cholesterol in human serum in two materials labelled as A and B measured in mg g$^{-1}$. We label the six data sets, two from each key comparison, as K2(Pb), K2(Cd), K5(N), K5(F), K6(A), and K6(B), respectively. These data are reproduced in appendix C. Since $s_b^2$(PM) is an optimal estimate of $\sigma_b^2$, we compare the estimates $s_b^2$(CA), $s_b^2$(DL), and $s_b^2$(C2) relative to $s_b^2$(PM) and the corresponding estimates $x_W$(CA), $x_W$(DL), and $x_W$(C2) relative to $x_W$(PM). Table 1 exhibits $s_b$(PM), $s_b$(CA), $s_b$(DL), and $s_b$(C2) for the six data sets. Table 2 exhibits $x_W$(PM), $x_W$(CA), $x_W$(DL), and $x_W$(C2) for the six data sets. For these six data sets, $s_b$(C2) is closer to the optimum value, $s_b$(PM), than $s_b$(DL). When $s_b$(CA) is zero, $s_b$(C2) is identical to $s_b$(DL). The combined result, $x_W$(C2), is closer to $x_W$(PM) than $x_W$(CA) and $x_W$(DL).

## 5. Summary

A classical statistics approach for combining the results from interlaboratory evaluations of a common measurand of value $\mu$ is to use a random effects model where the biases $(\mu_1 - \mu), \ldots, (\mu_m - \mu)$ in the laboratory results $x_1, \ldots, x_m$, respectively, are regarded as random variables having the same normal distribution with expected value zero and interlaboratory variance $\sigma_b^2$. The variances of $x_1, \ldots, x_m$ under the random effects model are $\sigma_b^2 + \sigma_1^2, \ldots, \sigma_b^2 + \sigma_m^2$, respectively, where $\sigma_1^2, \ldots, \sigma_m^2$ are intralaboratory variances. The most commonly used combined result is the weighted mean, $x_W$, of the individual results, $x_1, \ldots, x_m$, with weights proportional to their estimated variances under the random effects model. For simplicity of data analysis, many metrologists treat the estimated intralaboratory variances $s^2(x_1), \ldots, s^2(x_m)$ as the true intralaboratory variances, $\sigma_1^2, \ldots, \sigma_m^2$. Then determination of the combined result, $x_W$, reduces to the problem of estimating the interlaboratory variance, $\sigma_b^2$. We compared leading methods for estimating $\sigma_b^2$. The leading methods are those of Cochran, Paule and Mandel, and DerSimonian and Laird. We show that all three methods are special cases of a unifying identity. The unifying identity suggests a new two-step method. We used six data sets from three key comparisons for comparing the four methods for estimating $\sigma_b^2$. Rukhin *et al* show that the estimate of Paule and Mandel is optimal in the sense of being a conditionally REML estimate under normality, the condition being that the estimated variances $s^2(x_1), \ldots, s^2(x_m)$ be treated as the true variances, $\sigma_1^2, \ldots, \sigma_m^2$. So we compared the other three methods relative to the method of Paule and Mandel. The method of Paule and Mandel requires a simple iteration that can be easily done on a spreadsheet program. Therefore, it is the preferred method. If one must use a non-iterative method, the two-step method proposed here approximates the optimal method of Paule and Mandel better than the earlier methods of Cochran, and DerSimonian and Laird.

## Acknowledgments

I thank Andrew Rukhin for pointing me to the paper by DerSimonian and Laird and for his comments. Bert Rust, Geoffrey McFadden, Ron Boisvert, and the referee provided useful comments on an earlier draft of this paper.

## Appendix A

Since $E(x_i - x_C) = 0$ and

$$\begin{aligned} E(x_i - x_C)^2 &= V(x_i - x_C) \\ &= V(x_i) + V(x_C) - 2\mathrm{cov}(x_i, x_C) \\ &= V(x_i) + \frac{\sum_i a_i^2 V(x_i)}{[\sum_i a_i]^2} - \frac{2a_i V(x_i)}{\sum_i a_i}, \end{aligned}$$

we have

$$\begin{aligned} E\Big[\sum_i a_i (x_i - x_C)^2\Big] &= \sum_i a_i E(x_i - x_C)^2 \\ &= \sum_i a_i V(x_i) + \frac{\sum_i a_i^2 V(x_i)}{\sum_i a_i} - \frac{2\sum_i a_i^2 V(x_i)}{\sum_i a_i} \\ &= \sum_i a_i V(x_i) - \frac{\sum_i a_i^2 V(x_i)}{\sum_i a_i}. \end{aligned}$$

## Appendix B

For simplicity, write $y = \sigma_b^2$ in the function $F(\sigma_b^2)$ defined by equation (4). The function $F(y)$ is continuous. The first derivative is $\mathrm{d}F(y)/\mathrm{d}y = -1 \times \sum_i w_i^2 (x_i - x_W)^2$. This cannot be zero; otherwise $F(y)$ is not a function of $y$. So $\mathrm{d}F(y)/\mathrm{d}y$ is negative and $F(y)$ is strictly decreasing. The second derivative is $\mathrm{d}^2F(y)/\mathrm{d}y^2 = 2 \times [\sum_i w_i (z_i - z_W)^2]$,

**Table A1.** Key comparison CCQM K2: the measurands in the two data sets are, respectively, the amount of lead (Pb) and cadmium (Cd) in natural water measured in nmol $kg^{-1}$.

| NMI | Pb $x$ | Pb $s(x)$ | Cd $x$ | Cd $s(x)$ |
|---|---|---|---|---|
| PTB | 61.00 | 0.45 | 82.38 | 0.11 |
| NMi | 61.40 | 1.10 | 82.70 | 1.10 |
| NIMC | 62.21 | 0.30 | 82.90 | 0.63 |
| KRISS | 62.30 | 0.45 | 83.07 | 0.30 |
| LGC | 62.34 | 0.62 | 83.40 | 1.25 |
| NRC | 62.60 | 0.75 | 83.70 | 1.10 |
| IRMM | 62.70 | 0.26 | 83.90 | 0.90 |
| NIST | 62.84 | 0.15 | 84.60 | 1.00 |
| LNE | 65.90 | 1.35 | 84.80 | 1.95 |

**Table A2.** Key comparison CCQM K5: the measurands in the two data sets are, respectively, the mass fraction of pp′-DDE in natural (N) and fortified (F) fish oils measured in μg $g^{-1}$.

| NMI | N $x$ | N $s(x)$ | F $x$ | F $s(x)$ |
|---|---|---|---|---|
| BAM | 1.498 | 0.011 | 6.090 | 0.037 |
| KRISS | 1.525 | 0.006 | 6.001 | 0.012 |
| LGC | 1.554 | 0.012 | 5.989 | 0.111 |
| NARL | 1.493 | 0.032 | 5.905 | 0.066 |
| NIMC | 1.480 | 0.007 | 5.873 | 0.038 |
| NIST | 1.500 | 0.011 | 6.046 | 0.025 |
| NRC | 1.529 | 0.013 | 5.679 | 0.013 |
| NRCCRM | 1.481 | 0.008 | 6.035 | 0.022 |
| PTB | 1.535 | 0.008 | 6.037 | 0.033 |
| VNIIM | 1.606 | 0.007 | 6.301 | 0.032 |

**Table A3.** Key comparison CCQM K6: the measurands in the two data sets are, respectively, the mass fraction of cholesterol in human serum in material A and material B measured in mg $g^{-1}$.

| NMI | A $x$ | A $s(x)$ | B $x$ | B $s(x)$ |
|---|---|---|---|---|
| LGC | 2.214 | 0.0096 | 1.732 | 0.0066 |
| NARL | 2.250 | 0.0131 | 1.777 | 0.0170 |
| NIST | 2.215 | 0.0043 | 1.735 | 0.0033 |
| NMi–VSL | 2.137 | 0.0068 | 1.729 | 0.0045 |
| NMIJ | 2.195 | 0.0050 | 1.718 | 0.0039 |
| NRCCRM | 2.197 | 0.0062 | 1.736 | 0.0062 |
| PTB | 2.179 | 0.0114 | 1.705 | 0.0086 |

where $z_i = w_i(x_i - x_W)$ and $z_W = \sum_i w_i z_i / \sum_i w_i$. Since $\sum_i z_i = 0$, the second derivative cannot be zero; otherwise $F(y)$ is not a function of $y$. So $d^2F(y)/dy^2$ is positive and $F(y)$ is concave up. Thus, the maximum of $F(\sigma_b^2)$ occurs at $\sigma_b^2 = 0$ and $F(\sigma_b^2) \to -(m-1)$ as $\sigma_b^2 \to \infty$. When $F(0)$, i.e. the value of $F(\sigma_b^2)$ at $\sigma_b^2 = 0$, is positive, then by the intermediate value theorem of calculus a value of $\sigma_b^2$ exists for which $F(\sigma_b^2) = 0$. Since $F(\sigma_b^2)$ is strictly decreasing, such a value of $\sigma_b^2$ is unique. When $F(0)$ is negative, equation (4) has no positive solution. When $F(0)$ is zero, the solution is $\sigma_b^2 = 0$.

## Appendix C

In tables A1, A2 and A3, column 1 contains abbreviations of the participating national measurement institutes (NMIs), columns 2 and 3 contain the first data set, and columns 4 and 5 contain the second data set. In each data set, the result of measurement and its associated standard uncertainty are denoted by $x$ and $s(x)$, respectively. Source: http://kcdb.bipm.fr/BIPM-KCDB/.

## References


[1] Searle S R 1971 *Linear Models* (New York: Wiley)

[2] Kacker R N and Harville D A 1984 Approximations for standard errors of estimators of fixed and random effects in mixed linear models *J. Am. Stat. Assoc.* **79** 853–62

[3] Paule R C and Mandel J 1982 Consensus values and weighting factors *J. Res. Natl Bur. Stand.* **87** 377–85

[4] DerSimonian R and Laird N 1986 Meta-analysis in clinical trials *Control. Clin. Trials* **7** 177–87

[5] Cochran W G 1954 The combination of estimates from different experiments *Biometrics* **10** 101–29

[6] Schiller S B and Eberhardt K R1991 Combining data from independent chemical analysis methods *Spectrochim. Acta* B **46** 1607–13

[7] Rukhin A L, Biggerstaff B J and Vangel M G 2000 Restricted maximum likelihood estimation of common mean and the Mandel–Paule algorithm *J. Stat. Plann. Inf.* **83** 319–30

[8] Rukhin A L and Vangel M G 1998 Estimation of a common mean and weighted mean statistics *J. Am. Stat. Assoc.* **93** 303–8

[9] Harville D A 1977 Maximum likelihood approaches to variance component estimation and to related problems *J. Am. Stat. Assoc.* **72** 320–40

[10] Rukhin A L 2003 Two procedures of meta-analysis in clinical trials and interlaboratory studies (http://www.math.umbc.edu/~rukhin/papers/index.html)