Overall mean estimation of trace evidence in a two-level normal–normal model

Kool, Fréderique Suzanne; van Dorp, Inoni; Bolck, Annabel; Leegwater, Anna Jeannette ; Jongbloed, Geurt

DOI
10.1016/j.forsciint.2019.01.047

Publication date
2019

Document Version
Accepted author manuscript

Published in
Forensic Science International

Citation (APA)
Kool, F. S., van Dorp, I., Bolck, A., Leegwater, A. J., & Jongbloed, G. (2019). Overall mean estimation of trace evidence in a two-level normal–normal model. *Forensic Science International*, *297*, 342-349. https://doi.org/10.1016/j.forsciint.2019.01.047

Overall mean estimation of trace evidence in a two-level normal-normal model

Frederique Suzanne Kool, Inoni Nadine van Dorp, Annabel Bolck, Anna Jeanette Leegwater, Geurt Jongbloed
Postprint *Forensic Science International* **297** (2019), p 243-249

August 1, 2018

Abstract

In the evaluation of measurements on characteristics of forensic trace evidence, Aitken and Lucy (2004) model the data as a two-level model using assumptions of normality where likelihood ratios are used as a measure for the strength of evidence. A two-level model assumes two sources of variation: the variation within measurements in a group (first level) and the variation between different groups (second level). Estimates of the variation within groups, the variation between groups and the overall mean are required in this approach. This paper discusses three estimators for the overall mean. In forensic science, two of these estimators are known as the weighted and unweighted mean. For an optimal choice between these estimators, the within- and between-group covariance matrices are required. In this paper a generalization to the latter two mean estimators is suggested, which is referred to as the generalized weighted mean. The weights of this estimator can be chosen such that they minimize the variance of the generalized weighted mean. These optimal weights lead to a "toy estimator", because they depend on the unknown within- and between-group covariance matrices. Using these optimal weights with estimates for the within- and between-group covariance matrices leads to the third estimator, the optimal "plug-in" generalized weighted mean estimator. The three estimators and the toy estimator are compared through a simulation study. Under conditions generally encountered in practice, we show that the unweighted mean can be preferred over the weighted mean. Moreover, in these situations the unweighted mean and the optimal generalized weighted mean behave similarly. An artificial choice of parameters is used to provide an example where the optimal generalized weighted mean outperforms both the weighted and unweighted mean. Finally, the three mean estimators are applied to real XTC data to illustrate the impact of the choice of overall mean estimator.

1 Introduction

The likelihood ratio is a generally accepted measure for the strength of evidence in many forensic comparison problems. Modelling the data as a two-level random effects model using assumptions of normality is a well-known approach in likelihood ratio calculation [1, 6]. The use of a two-level model leads to a likelihood ratio which depends on the unknown parameters of the two-level model. Within the Likelihood Paradigm [11] estimates of these parameters are required to estimate the likelihood ratio. Alternatively, it is possible to assign priors to all parameters following a full-Bayesian approach [3, 5, 14]. In this paper, different methods are described to estimate one of the parameters: the overall mean vector of the two-level model. Two currently used estimators in forensic statistics, the weighted and unweighted mean, are compared. There is still discussion which of these mean estimators should be used when the data are unbalanced, i.e. when the number of data points differs per group [1, 12]. Moreover, a general class of estimators for the overall mean, referred to as generalized weighted mean, is suggested. This class contains the two aforementioned estimators as special cases. The choice of the mean estimator is important for the

commonly used analysis of variance estimator to estimate the between-source covariance matrix, which is another parameter to be estimated in the two-level model [12].

In Section 2 the likelihood ratio approach in the setting of a two-level model is described, yielding an explicit expression for the likelihood ratio in terms of the model parameters. Section 3 covers the explanation of the estimators and their relative efficiencies in terms of (partly) unknown parameters. In Section 4 a comparison of the estimation techniques is given through a simulation study and in Section 5 the estimators are applied to real XTC data. In this paper the results are given for the multivariate case. The results for the univariate case are obtained by replacing the (traces of the) covariance matrices with the corresponding variances.

2 Likelihood ratio approach

In forensic comparison problems it is investigated whether a control item (e.g. XTC tablets from consignment C_1) and a recovered item (e.g. XTC tablets from consignment C_2) originate from the same unknown source[1]. Very generally stated, a prosecutor's hypothesis (H_p) and a hypothesis of the defence (H_d) may be as follows:

$$\begin{cases} H_p: & \text{The control and recovered item originate from the same source.} \\ H_d: & \text{The control and recovered item originate from different sources.} \end{cases}$$

Comparison of the control and recovered item given the two hypotheses involves evidence E. This evidence concerns certain characteristics or features of the two items. The likelihood ratio approach refers to a well-known probabilistic framework based on Bayes' rule to evaluate the strength of the evidence in such forensic comparison problems. In this approach, the likelihood ratio is the ratio of the probability of evidence E given the two hypotheses H_p and H_d:

$$\text{LR} = \frac{P(E \mid H_p)}{P(E \mid H_d)}. \tag{1}$$

This likelihood ratio expresses how much more likely it is to find the evidence under the prosecutor's hypothesis than under the hypothesis of the defence. Therefore, the likelihood ratio can be seen as a measure to quantify the strength of evidence.

2.1 Model

Various types of models exist to compute the likelihood ratio in equation (1). In this paper, the focus will be on a feature-based two-level random effects model using assumptions of normality which is applicable to continuous data [2, 6].

Consider the situation that several continuous features of the control and recovered item are measured by forensic experts, e.g. the diameter, thickness and weight of the XTC tablets in consignment C_1 and C_2. Let k denote the number of features and let n_1 be the number of measurements of these features on the control item, e.g. the number of tablets that is measured in consignment C_1. The composed continuous random vector $\mathbf{Y}_1$ represents the n_1 measurement vectors of the features on the control item,

$$\mathbf{Y}_1 = (\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1}) = \left(\begin{bmatrix} Y_{11,1} \\ Y_{11,2} \\ \vdots \\ Y_{11,k} \end{bmatrix}, \ldots, \begin{bmatrix} Y_{1n_1,1} \\ Y_{1n_1,2} \\ \vdots \\ Y_{1n_1,k} \end{bmatrix} \right).$$

[1] In the context of [8], this problem is known as a common source problem. The model corresponds to the situation where the sources are assumed to be random realizations from a probability distribution. For more details about the difference between common and specific source problems, see [9].

This vector can be referred to as control data. The control data will be compared to the recovered data $\mathbf{Y}_2$, i.e. the composed random vector which represents the n_2 measurements of the features on the recovered item. Thus, the composed random vectors $\mathbf{Y}_l = (\mathbf{Y}_{lj}, 1 \leq j \leq n_l), l = 1, 2$, represent for example the diameters, thicknesses and weights of the tablets from consignments C_1 and C_2. To compare the control and recovered item, the means of the control and recovered data can be used as the evidence, i.e.

$$E = (\overline{\mathbf{Y}}_1, \overline{\mathbf{Y}}_2)$$

where

$$\overline{\mathbf{Y}}_l = \frac{1}{n_l} \sum_{j=1}^{n_l} \mathbf{Y}_{lj} \qquad \text{for } l = 1, 2$$

denotes the mean over the n_l measurements.

The data are modelled using a (two-level) random effects model under the assumption of normality [2, 6]. The use of such a two-level model is appropriate, because the data are organized at more than one level: the measurements (first level) are nested within the items (second level), such as the control and recovered item. The variation between the n_l measurements within the same item is known as the within-source variation. The variation between the items is known as the between-source variation. It is assumed that both the within- and between-source variation are multivariate normally distributed. This means that within a source, the control and recovered data are independent and normally distributed around their group means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, i.e.

$$\overline{\mathbf{Y}}_l \mid \boldsymbol{\theta}_l \sim \mathcal{N}_k \left(\boldsymbol{\theta}_l, n_l^{-1}\boldsymbol{\Sigma}\right) \qquad \text{for } l = 1, 2$$

and the between-source variation is modelled by independent normally distributed random variables

$$\boldsymbol{\theta}_l \sim \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{T}) \qquad \text{for } l = 1, 2.$$

2.2 Likelihood ratio

In the literature, explicit likelihood ratio formulas under the normality assumptions in the two-level model are derived [1, 6, 15]. In this paper we will use the following likelihood ratio of the observed evidence $E = (\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2)$ [4]:

$$\mathrm{LR}(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2 | \boldsymbol{\mu}) = \frac{|\mathbf{U}_0|^{\frac{1}{2}}}{|\mathbf{U}_n|^{\frac{1}{2}}} \exp\left[\frac{1}{2}\left((\overline{\mathbf{y}}_2 - \boldsymbol{\mu})^T \mathbf{U}_0^{-1} (\overline{\mathbf{y}}_2 - \boldsymbol{\mu}) - (\overline{\mathbf{y}}_2 - \boldsymbol{\mu}_n)^T \mathbf{U}_n^{-1} (\overline{\mathbf{y}}_2 - \boldsymbol{\mu}_n)\right)\right] \tag{2}$$

where

$$\begin{aligned}
\mathbf{U}_0 &= \mathbf{T} + n_2^{-1}\boldsymbol{\Sigma}, \\
\mathbf{U}_n &= \mathbf{T}_n + n_2^{-1}\boldsymbol{\Sigma}, \\
\boldsymbol{\mu}_n &= \mathbf{T}(\mathbf{T} + n_1^{-1}\boldsymbol{\Sigma})^{-1}\overline{\mathbf{y}}_1 + n_1^{-1}\boldsymbol{\Sigma}(\mathbf{T} + n_1^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{\mu}, \\
\mathbf{T}_n &= \mathbf{T} - \mathbf{T}(\mathbf{T} + n_1^{-1}\boldsymbol{\Sigma})^{-1}\mathbf{T}.
\end{aligned}$$

The explicit likelihood ratio formulas depend on the unknown overall mean $\boldsymbol{\mu}$, the between-source covariance matrix $\mathbf{T}$ and the within-source covariance matrix $\boldsymbol{\Sigma}$ of the described two-level model. Hence, in the Likelihood Paradigm, estimates of these parameters are required to estimate the likelihood ratio. In Section 3, estimators for the overall mean $\boldsymbol{\mu}$ are described. Estimators for the covariance matrices $\mathbf{T}$ and $\boldsymbol{\Sigma}$ are for example the multivariate analysis of variance estimators [12, 13]. Next to the computation of the likelihood ratio, the choice of the mean estimator $\hat{\boldsymbol{\mu}}$ is important for the analysis of variance estimator of the between-source covariance matrix $\mathbf{T}$, because this quantity depends on the mean $\boldsymbol{\mu}$ [12]. As an alternative to these approaches, maximum

likelihood estimators can be used [13]. However, in the two-level normal-normal setup no explicit formulas exist for these estimators. Therefore, iterative methods are required [12]. Another option is to use a full-Bayesian approach with priors assigned to all parameters [3, 5, 14]. In this paper, we will focus on the non-Bayesian approach with $\mathbf{\Sigma}$ and $\mathbf{T}$ fixed, and we will compare several estimators for $\boldsymbol{\mu}$.

2.3 Background data

To estimate the parameters of the two-level model, background data that represent the population are required. The background data consist of measurements of the continuous features on a random sample of m items or groups, which represent the population. In each of the m groups, n_i $(i = 1, \ldots, m)$ measurements are taken. The background data are denoted as $\{\mathbf{Z}_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n_i\}$, where $\mathbf{Z}_{ij}$ represents the vector of measured features within group i of measurement j. The background data are modelled by the extension of the two-level model described in Section 2.2 [1],

$$\begin{aligned} \mathbf{Z}_{ij} \mid \boldsymbol{\theta}_i &\overset{\text{iid}}{\sim} \mathcal{N}_k(\boldsymbol{\theta}_i, \mathbf{\Sigma}), \quad 1 \leq j \leq n_i, \\ \boldsymbol{\theta}_i &\overset{\text{iid}}{\sim} \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{T}), \quad 1 \leq i \leq m. \end{aligned}$$

Under these assumptions, the background data are in fact modelled by a random effects model [13], i.e.

$$\mathbf{Z}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{ij} \qquad \text{for } 1 \leq i \leq m, \quad 1 \leq j \leq n_i,$$

with $\boldsymbol{\mu}$ the overall mean,

$$\boldsymbol{\alpha}_i \overset{\text{iid}}{\sim} \mathcal{N}_k(\mathbf{0}_k, \mathbf{T}), \qquad 1 \leq i \leq m,$$

the random group effect and, independent of the $\boldsymbol{\alpha}_i$'s,

$$\boldsymbol{\varepsilon}_{ij} \overset{\text{iid}}{\sim} \mathcal{N}_k(\mathbf{0}_k, \mathbf{\Sigma}), \qquad 1 \leq j \leq n_i,$$

the random noise vectors or within-source variation.

3 Estimating the overall mean

First, the weighted mean and the unweighted mean are discussed as estimators for the overall mean $\boldsymbol{\mu}$. In Section 3.2 it is shown that what is the best estimator (the estimator with smallest variance) depends on the ratio of the traces of the within- and between-source covariance matrices. To derive this, a multivariate generalization of variance is given in Section 3.1. In Section 3.3 a generalization of the weighted and unweighted mean estimators is suggested, which is referred to as the generalized weighted mean. The weights of this estimator can be chosen such that they minimize the variance of the generalized weighted mean. These optimal weights lead to what we will call a "toy estimator". We use the term "toy estimator", because the optimal weights depend on the unknown within- and between-source covariance matrices $\mathbf{\Sigma}$ and $\mathbf{T}$. Hence, in practice only an estimate of the optimal weights can be obtained and the resulting estimator will be referred to as the optimal "plug-in" generalized weighted mean estimator.

3.1 Multivariate generalization of variance

A natural choice for the multivariate concept of variance for unbiased estimators is to consider the expected value of the squared Euclidean distance between the estimator and the true parameter of interest, i.e.

$$\text{Var}(\hat{\boldsymbol{\mu}}) := \text{E}\left[||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2\right].$$

Note that we will prefer the unbiased estimator with minimal expected distance to the true parameter. For unbiased estimators it follows that

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\mu}}) &= \mathrm{E}\left[||\hat{\boldsymbol{\mu}} - \mathrm{E}[\hat{\boldsymbol{\mu}}]||^2\right] = \mathrm{E}\left[(\hat{\boldsymbol{\mu}} - \mathrm{E}[\hat{\boldsymbol{\mu}}])^T(\hat{\boldsymbol{\mu}} - \mathrm{E}[\hat{\boldsymbol{\mu}}])\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{k}(\hat{\mu}_i - \mathrm{E}[\hat{\mu}_i])^2\right] = \sum_{i=1}^{k}\mathrm{Var}(\hat{\mu}_i) = tr(\boldsymbol{\Sigma}),\end{aligned}$$

where $\boldsymbol{\Sigma}$ denotes the covariance matrix of $\hat{\boldsymbol{\mu}}$. Any further mention of variance will refer to this definition.

3.2 Weighted versus unweighted mean

The group means of the background data are defined as the average of the observations $\mathbf{Z}_{ij}$ in each group,

$$\overline{\mathbf{Z}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{Z}_{ij}, \qquad 1 \leq i \leq m, \tag{3}$$

such that $\overline{\mathbf{Z}}_i \sim \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})$. These group means are used to approximate $\boldsymbol{\theta}_i$. Two estimators for the overall mean $\boldsymbol{\mu}$ are the weighted mean and the unweighted mean. The weighted mean is the average over all observations $\mathbf{Z}_{ij}$ in the background data [13],

$$\hat{\boldsymbol{\mu}}_w = \frac{1}{N}\sum_{i=1}^{m} n_i\overline{\mathbf{Z}}_i = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathbf{Z}_{ij}, \tag{4}$$

where N is the total number of observations, i.e. $N = \sum_{i=1}^{m} n_i$. The weighted mean is unbiased, since

$$\mathrm{E}[\hat{\boldsymbol{\mu}}_w] = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathrm{E}\left[\mathbf{Z}_{ij}\right] = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathrm{E}\left[\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{ij}\right] = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{\mu} = \boldsymbol{\mu}.$$

The variance of the weighted mean is equal to

$$\mathrm{Var}(\hat{\boldsymbol{\mu}}_w) = \frac{tr(\mathbf{T})}{N^2}\sum_{i=1}^{m} n_i^2 + \frac{tr(\boldsymbol{\Sigma})}{N},$$

see Appendix A.1. The unweighted mean is the mean of the group means [12],

$$\hat{\boldsymbol{\mu}}_u = \frac{1}{m}\sum_{i=1}^{m}\overline{\mathbf{Z}}_i. \tag{5}$$

The unweighted mean is also unbiased, since

$$\mathrm{E}[\hat{\boldsymbol{\mu}}_u] = \frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left[\overline{\mathbf{Z}}_i\right] = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{\mu} = \boldsymbol{\mu}$$

and its variance is equal to

$$\mathrm{Var}(\hat{\boldsymbol{\mu}}_u) = \frac{tr(\mathbf{T})}{m} + \frac{tr(\boldsymbol{\Sigma})}{m^2}\sum_{i=1}^{m}\frac{1}{n_i},$$

see Appendix A.1.

First note that if the data are balanced, i.e. $n_i = n$ for all $i = 1, \ldots, m$, the weighted and unweighted mean are exactly the same. For unbalanced data where the number of measurements

differs per group, there is a dispute whether to use the weighted mean or the unweighted mean [1, 12]. The weighted mean fits naturally with a designed experiment or other reasons where the unequal number of measurements reflects the composition of the population or the importance of the groups. In that case it is important that groups with more measurements have more weight in the estimation of the overall mean, which is an argument in favor of the weighted mean. In cases where the number of measurements is more or less randomly chosen or determined by factors independent of the population composition (e.g. sampling costs) the number of measurements is not important. It is then beneficial that groups have equal importance, despite the number of observations, which is an argument in favor of the unweighted mean. In fact, below it is shown that the best choice between these estimators depends on the situation.

Since both estimators are unbiased, it will be examined which estimator has smallest variance. Hence, consider the efficiency of $\hat{\boldsymbol{\mu}}_w$ relative to $\hat{\boldsymbol{\mu}}_u$ [10]:

$$\text{eff}(\hat{\boldsymbol{\mu}}_u, \hat{\boldsymbol{\mu}}_w) = \frac{\text{Var}(\hat{\boldsymbol{\mu}}_w)}{\text{Var}(\hat{\boldsymbol{\mu}}_u)} = \frac{\frac{tr(\mathbf{T})}{N^2}\sum_{i=1}^m n_i^2 + \frac{tr(\boldsymbol{\Sigma})}{N}}{\frac{tr(\mathbf{T})}{m} + \frac{tr(\boldsymbol{\Sigma})}{m^2}\sum_{i=1}^m \frac{1}{n_i}}. \tag{6}$$

Multiplying the numerator and denominator in equation (6) with the term m^2N^2 and setting $r = \frac{tr(\boldsymbol{\Sigma})}{tr(\mathbf{T})}$ results in

$$\text{eff}(\hat{\boldsymbol{\mu}}_u, \hat{\boldsymbol{\mu}}_w) = \frac{m^2\sum_{i=1}^m n_i^2 + m^2Nr}{mN^2 + rN^2\sum_{i=1}^m \frac{1}{n_i}}. \tag{7}$$

Using Jensen's inequality it can be shown that the efficiency can have larger and smaller values than one, see Appendix A.2. Therefore, one cannot be conclusive about which estimator has smallest variance. From Appendix A.2, it follows that

$$\text{eff}(\hat{\boldsymbol{\mu}}_u, \hat{\boldsymbol{\mu}}_w) > 1 \quad \text{iff} \quad r < \frac{m^2\sum_{i=1}^m n_i^2 - mN^2}{N^2\sum_{i=1}^m \frac{1}{n_i} - m^2N} =: c. \tag{8}$$

Note that both the numerator and the denominator of c are positive because of inequalities (17) and (18) (see Appendix A.2), hence the constant c is always positive. Therefore,

$$\begin{cases} \text{Var}(\hat{\boldsymbol{\mu}}_w) > \text{Var}(\hat{\boldsymbol{\mu}}_u) & \text{if } tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T}), \\ \text{Var}(\hat{\boldsymbol{\mu}}_w) < \text{Var}(\hat{\boldsymbol{\mu}}_u) & \text{if } tr(\boldsymbol{\Sigma}) > c \cdot tr(\mathbf{T}). \end{cases} \tag{9}$$

From the inequalities in (9) it follows that the best choice of the estimator depends on two factors. The first is the ratio between the trace of the within-source covariance matrix $\boldsymbol{\Sigma}$ and the trace of the between-source covariance matrix $\mathbf{T}$. For example, if the trace of the within-source covariance matrix $\boldsymbol{\Sigma}$ is small, i.e. $\overline{\mathbf{Z}}_i \approx \boldsymbol{\theta}_i$, the unweighted mean virtually equals the maximum likelihood estimator based on the (unobservable) $\boldsymbol{\theta}_i$'s and we would prefer the unweighted mean. This example corresponds to the first inequality in expression (9). Since the parameters $\boldsymbol{\Sigma}$ and $\mathbf{T}$ are unknown, this factor relies on prior knowledge or on experience of the forensic expert. The second factor that affects the choice between the weighted and unweighted mean is the value of the constant c, which depends on the number of groups m and the number of measurements within each group n_i. The following lemma gives more insight in the possible values of the constant c.

Lemma 3.1. *The constant*

$$c = \frac{m^2\sum_{i=1}^m n_i^2 - mN^2}{N^2\sum_{i=1}^m \frac{1}{n_i} - m^2N}$$

is always greater than or equal to 1.

The proof of this lemma can be found in Appendix A.3. This lemma illustrates that when $tr(\boldsymbol{\Sigma}) < tr(\mathbf{T})$ the unweighted mean will always have a smaller variance than the weighted mean.

In many forensic comparison problems it is realistic to assume that within-source variation is smaller than between-source variation. For instance, in XTC comparison problems this is due to the fact that the errors that cause the within-group variation (e.g. measurement errors, production errors, inhomogeneity within a batch) are often smaller than the between-group variation (mainly based on the preference of the producers). Consequently, in many XTC comparison problems it can be assumed that the trace of the within-source covariance matrix $\mathbf{\Sigma}$ is smaller than the trace of the between-source covariance matrix $\mathbf{T}$, i.e. $tr(\mathbf{\Sigma}) < tr(\mathbf{T})$. Since $c \geq 1$ always holds, the unweighted mean should in these situations be preferred over the weighted mean.

3.3 Generalized weighted mean

This section suggests a more general estimator for the mean compared to the weighted and unweighted mean described in Section 3.2. This general estimator will be referred to as the generalized weighted mean[2]. Define the generalized weighted mean as [10]

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{m} \mathbf{W}_i \overline{\mathbf{Z}}_i \qquad \text{where } \mathbf{W}_i \text{ is a } k \times k \text{ matrix such that } \sum_{i=1}^{m} \mathbf{W}_i = \mathbf{I}_k. \tag{10}$$

Here, $\mathbf{I}_k$ denotes the $k \times k$-dimensional identity matrix. The weighted and unweighted mean are special cases of the generalized weighted mean given in equation (10). It can be seen that the weighted mean $\hat{\boldsymbol{\mu}}_w$ is the generalized weighted mean with weight matrices $\mathbf{W}_i = \frac{n_i}{N}\mathbf{I}_k$ for $1 \leq i \leq m$. The unweighted mean $\hat{\boldsymbol{\mu}}_u$ is the generalized weighted mean with weight matrices $\mathbf{W}_i = \frac{1}{m}\mathbf{I}_k$ for $1 \leq i \leq m$.

Since the weight matrices $\mathbf{W}_1, \ldots, \mathbf{W}_m$ in equation (10) add up to the identity matrix, it follows that the generalized weighted mean is unbiased, i.e.

$$\mathrm{E}(\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{m} \mathbf{W}_i \,\mathrm{E}\left(\overline{\mathbf{Z}}_i\right) = \left(\sum_{i=1}^{m} \mathbf{W}_i\right) \boldsymbol{\mu} = \boldsymbol{\mu}.$$

The covariance matrix of $\hat{\boldsymbol{\mu}}$ is equal to

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{m} \mathbf{W}_i \,\mathrm{Cov}\left(\overline{\mathbf{Z}}_i, \overline{\mathbf{Z}}_i\right) \mathbf{W}_i^T = \sum_{i=1}^{m} \mathbf{W}_i (\mathbf{T} + n_i^{-1}\mathbf{\Sigma}) \mathbf{W}_i^T \tag{11}$$

so that its variance is given by

$$\mathrm{Var}(\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{m} tr\left(\mathbf{W}_i (\mathbf{T} + n_i^{-1}\mathbf{\Sigma}) \mathbf{W}_i^T\right).$$

Since the variance depends on the choice of the weight matrices $\mathbf{W}_1, \ldots, \mathbf{W}_m$, the question arises how to choose these weights to minimize $\mathrm{Var}(\hat{\boldsymbol{\mu}})$ subject to the constraint $\sum_{i=1}^{m} \mathbf{W}_i = \mathbf{I}_k$.[3]

[2] In the literature this estimator is called the weighted mean. However, in forensic literature the estimator in equation (4) is called the weighted mean. Therefore, we will refer to this estimator as generalized weighted mean.

[3] If only diagonal matrices would be considered, a similar analysis shows that the matrix with weights $\mathbf{w}_i = (w_{i1}, \ldots, w_{ik})^T$ on the diagonal that minimizes $\mathrm{Var}(\hat{\boldsymbol{\mu}})$ subject to the constraint $\sum_{i=1}^{m} \mathbf{w}_i = \mathbf{1}_k$ is found from

$$\mathbf{w}_i = \left(\sum_{j=1}^{m} \frac{1}{\mathrm{diag}(\mathbf{T} + n_j^{-1}\mathbf{\Sigma})}\right)^{-1} \frac{1}{\mathrm{diag}(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})}, \qquad 1 \leq i \leq m,$$

where all vector products and divisions are elementwise. Choosing the diagonal matrix with these weights results in a better mean estimator in terms of variance than $\hat{\boldsymbol{\mu}}_w$ and $\hat{\boldsymbol{\mu}}_u$, but it will not be as good as $\hat{\boldsymbol{\mu}}_{\mathrm{opt}}$, which is the optimal mean estimator.

Lemma 3.2. *The weights* $\mathbf{W}_1, \ldots, \mathbf{W}_m$ *that minimize* $\mathrm{Var}(\hat{\boldsymbol{\mu}})$ *subject to the constraint* $\sum_{i=1}^{m} \mathbf{W}_i = \mathbf{I}_k$ *are given by*

$$\mathbf{W}_i = \left(\sum_{j=1}^{m} (\mathbf{T} + n_j^{-1}\boldsymbol{\Sigma})^{-1} \right)^{-1} (\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})^{-1} \tag{12}$$

where $i = 1, \ldots, m$.

The proof of this lemma is given in Appendix A.4. This lemma shows that the weights in equation (12) minimize the variance of the generalized weighted mean. Hence, these optimal weights lead to the following "toy estimator":

$$\hat{\boldsymbol{\mu}}_{\text{opt}} = \left(\sum_{i=1}^{m} (\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})^{-1} \right)^{-1} \left(\sum_{i=1}^{m} (\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})^{-1} \overline{\mathbf{Z}}_i \right). \tag{13}$$

Since the weights in equation (12) yield minimum variance for $\hat{\boldsymbol{\mu}}$ we can thus conclude that, if the parameters $\boldsymbol{\Sigma}$ and $\mathbf{T}$ are known, $\hat{\boldsymbol{\mu}}_{\text{opt}}$ is the best of these three estimators.

However, in practice this result is not immediately useful because the optimal weights depend on the unknown parameters $\boldsymbol{\Sigma}$ and $\mathbf{T}$. If estimated values for these parameters are substituted in the optimal weights, this will influence the variance of the toy estimator in equation (13) and the resulting estimator will be biased. For example, the multivariate analysis of variance estimators [12] for $\boldsymbol{\Sigma}$ and $\mathbf{T}$ could be used, which are given by

$$\begin{aligned}
&\hat{\boldsymbol{\Sigma}} = \frac{1}{N-m} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\mathbf{z}_{ij} - \overline{\mathbf{z}}_{i\cdot})(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{i\cdot})^T \qquad \text{where } \overline{\mathbf{z}}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{z}_{ij}, \\
&\hat{\mathbf{T}} = \frac{\mathrm{MS}^2_{\text{between}} - \hat{\boldsymbol{\Sigma}}}{\kappa} \qquad \text{where } \kappa = \frac{1}{m-1} \left(N - \frac{\sum_{i=1}^{m} n_i^2}{N} \right), \\
&\mathrm{MS}^2_{\text{between}} = \frac{1}{m-1} \sum_{i=1}^{m} n_i (\overline{\mathbf{z}}_{i\cdot} - \overline{\mathbf{z}})(\overline{\mathbf{z}}_{i\cdot} - \overline{\mathbf{z}})^T \text{ and } \overline{\mathbf{z}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \mathbf{z}_{ij}.
\end{aligned} \tag{14}$$

The performance of the plug-in estimator $\hat{\boldsymbol{\mu}}_{\text{plug}}$ based on these estimates for $\boldsymbol{\Sigma}$ and $\mathbf{T}$ will be further evaluated in the following sections. Introducing the toy estimator gives more theoretical insight in the various estimators for the overall mean $\boldsymbol{\mu}$. In the results of the simulation study in Section 4 this will be further explored.

4 Simulation study

In this section the mean estimators of Section 3 are compared in a simulation study. In Section 4.1 the performance of the weighted and unweighted mean estimators is compared using Monte Carlo simulation. In Section 4.2, this comparison is extended with the optimal generalized weighted mean estimator. Since this is a toy estimator and cannot be computed in practice, the optimal generalized weighted mean with estimates for the within- and between-source covariance matrices will also be considered, which will be called the optimal "plug-in" generalized weighted mean estimator. Finally, in Section 4.3 an artificial choice of parameters is used to show some examples where the optimal generalized weighted mean outperforms both the weighted and unweighted mean.

4.1 Weighted versus unweighted mean

In expression (9) we have seen that the best choice between the weighted and unweighted mean depends on the ratio of the traces of the within- and between-source covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{T}$.

However, the covariance matrices $\mathbf{\Sigma}$ and $\mathbf{T}$ are unknown. Hence, to use expression (9) in practice, one should have prior knowledge about the ratio between $tr(\mathbf{\Sigma})$ and $tr(\mathbf{T})$. In many comparison problems the trace of the within-source covariance matrix can be assumed to be smaller than the trace of the between-source covariance matrix. Furthermore, in Lemma 3.1 it is shown that the value of the constant c will always be larger than one. Therefore, it is expected that in most cases the unweighted mean has a smaller variance than the weighted mean.

Given this result, it is interesting to compare the performance of the weighted and the unweighted mean in estimating the true mean $\boldsymbol{\mu}$. To this end, we perform two Monte Carlo simulations. In these simulations, the values for the number of groups m are set to $m = 10$ and $m = 1200$, respectively, and the number of measurements in each group $n_i, 1 \leq i \leq m$, is drawn randomly, where values $1 \leq n_i \leq 20$ are used. Given these values of n_i and m, a background data set is generated M times according to the model described in Section 2.3. To simulate the background data set in both situations, the parameters $\boldsymbol{\mu}$, $\mathbf{\Sigma}$ and $\mathbf{T}$ are fixed based on diameter (in millimeters), thickness (in millimeters) and weight (in milligrams) observations in real XTC tablet comparisons. These values are given by:

$$\begin{gathered}\boldsymbol{\mu} = \begin{bmatrix} 8.242 \\ 4.528 \\ 276.0 \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} 0.002013 & 0.0007271 & 0.01408 \\ 0.0007271 & 0.03046 & 0.6133 \\ 0.01408 & 0.6133 & 90.61 \end{bmatrix}, \\ \mathbf{T} = \begin{bmatrix} 0.6026 & 0.06689 & 31.56 \\ 0.06689 & 0.6371 & 32.90 \\ 31.56 & 32.90 & 3562 \end{bmatrix}.\end{gathered} \tag{15}$$

The results of the two Monte Carlo simulations are given for each element of the three-dimensional estimated mean vector and can be found in the box plots in Figure 1. From these figures it can be seen that the estimated values of the two mean estimators are close. As can be expected, if there are more observations in the background data (1200 groups), the estimates are more accurate compared to the estimates using fewer observations (10 groups).

The mean squared error (MSE) [10] is chosen as a measure of performance for the estimators. The MSE measures the average of the squared values of the errors, i.e. the Euclidean distance between the estimate and the true value $\boldsymbol{\mu}$:

$$\mathrm{E}\left[||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2\right].$$

Hence, a mean squared error of zero means that the estimator estimates the true mean $\boldsymbol{\mu}$ perfectly. The estimators can be compared by using their MSEs, where the smallest MSE is preferred. For the unbiased weighted and unweighted mean, the MSE equals the variance of the estimators. Hence, minimizing the mean squared error is equivalent to minimizing the variance and the estimators with the lowest MSE are thus the most efficient.

To compute the MSE based on the Monte Carlo simulation, for each simulation i, with $1 \leq i \leq M$, the squared Euclidean distance between the estimate and the true value is computed. After M simulations the average over these squared distances is taken as the (numerically approximated) mean squared error. The resulting mean squared errors are given in Table 1.

From these MSE values it is clear that the performance of the estimators increases when the number of groups m is higher. Since the MSEs of the unweighted mean are smaller than the MSEs of the weighted mean, the unweighted mean should be preferred over the weighted mean. For both simulations the constant c can be computed and equals $c = 4.48$ for 10 groups and $c = 3.43$ for 1200 groups and with $tr(\mathbf{\Sigma}) = 90.6$ and $tr(\mathbf{T}) = 3563$ it can be seen that $tr(\mathbf{\Sigma}) < c \cdot tr(\mathbf{T})$. Consequently, from expression (9) it follows that the variance of the unweighted mean is smaller than the variance of the weighted mean.

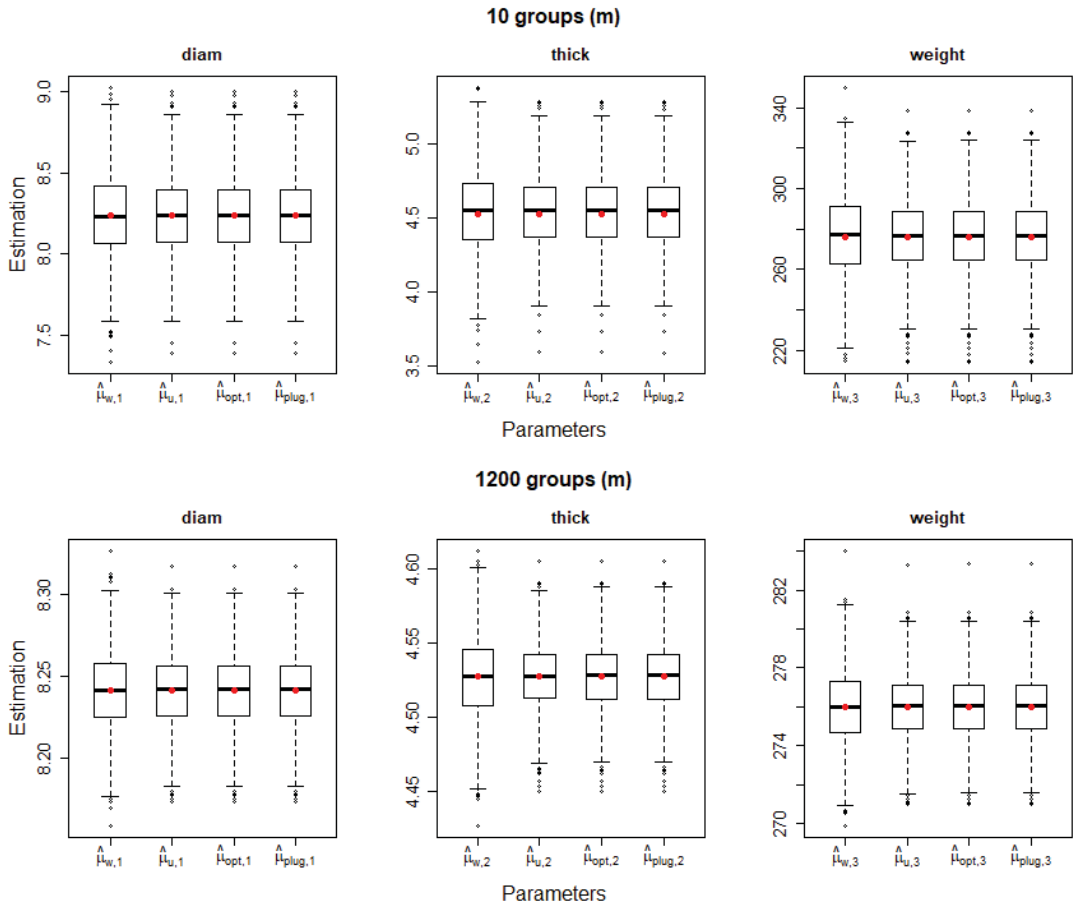

Figure 1: Box plots of estimated values from $\hat{\boldsymbol{\mu}}_w$, $\hat{\boldsymbol{\mu}}_u$, $\hat{\boldsymbol{\mu}}_{\text{opt}}$ and $\hat{\boldsymbol{\mu}}_{\text{plug}}$ for two Monte Carlo simulations ($M = 1000$) with parameters given as in equation (15). The red dot indicates the true overall mean value.

m	MSE $\hat{\boldsymbol{\mu}}_w$	MSE $\hat{\boldsymbol{\mu}}_u$	MSE $\hat{\boldsymbol{\mu}}_{\text{opt}}$	MSE $\hat{\boldsymbol{\mu}}_{\text{plug}}$
10	406	352	352	352
1200	3.93	2.96	2.96	2.96

Table 1: Mean squared errors of the estimated mean using the weighted mean $\hat{\boldsymbol{\mu}}_w$, the unweighted mean $\hat{\boldsymbol{\mu}}_u$, the toy estimator $\hat{\boldsymbol{\mu}}_{\text{opt}}$ and the plug-in estimator $\hat{\boldsymbol{\mu}}_{\text{plug}}$ for two Monte Carlo simulations ($M = 1000$) with parameters as given in equation (15).

Since the values for the overall mean $\boldsymbol{\mu}$ and the covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{T}$ are fixed, it is possible to determine the true value of the likelihood ratio for this problem. Therefore, five measurements for both the control and recovered data are generated, assuming that the prosecutor's hypothesis is true, i.e. that the control and recovered item originate from the same source. Using equation (2) with the parameters given in equation (15), the true value of the likelihood ratio is found. Keeping the covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{T}$ fixed, the likelihood ratios based on $\hat{\boldsymbol{\mu}}_w$ and $\hat{\boldsymbol{\mu}}_u$ can also be calculated. The mean squared error for the likelihood ratio values is then computed

by

$$\frac{1}{M}\sum_{i=1}^{M}\left[\mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2|\boldsymbol{\mu}) - \mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2|\hat{\boldsymbol{\mu}}(i))\right]^2$$

for each Monte Carlo simulation i, with $1 \leq i \leq M$. The resulting mean squared errors can be found in Table 2.

Clearly, the MSE values of the likelihood ratios reduce significantly when the number of groups m is higher. Moreover, the performance of the unweighted mean is significantly better than the performance of the weighted mean. Combining this observation with the fact that the unweighted mean is more efficient than the weighted mean, the unweighted mean should in this situation be preferred over the weighted mean.

m	MSE $\mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2\|\hat{\boldsymbol{\mu}}_w)$	MSE $\mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2\|\hat{\boldsymbol{\mu}}_u)$	MSE $\mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2\|\hat{\boldsymbol{\mu}}_{\mathrm{opt}})$	MSE $\mathrm{LR}(\overline{\mathbf{y}}_1,\overline{\mathbf{y}}_2\|\hat{\boldsymbol{\mu}}_{\mathrm{plug}})$
10	$3.97 \cdot 10^6$	$2.61 \cdot 10^6$	$2.60 \cdot 10^6$	$2.61 \cdot 10^6$
1200	$6.34 \cdot 10^3$	$4.99 \cdot 10^3$	$4.96 \cdot 10^3$	$4.96 \cdot 10^3$

Table 2: Mean squared errors of the estimated likelihood ratio using the weighted mean $\hat{\boldsymbol{\mu}}_w$, the unweighted mean $\hat{\boldsymbol{\mu}}_u$, the toy estimator $\hat{\boldsymbol{\mu}}_{\mathrm{opt}}$ and the plug-in estimator $\hat{\boldsymbol{\mu}}_{\mathrm{plug}}$ for two Monte Carlo simulations ($M = 1000$) with parameters as given in equation (15). The true likelihood ratio is equal to $1.11 \cdot 10^3$.

4.2 Generalized weighted mean

In the Monte Carlo simulations in Section 4.1 the values for the covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{T}$ are fixed, see equation (15). Substituting these values into the toy estimator in equation (13), the toy estimator yields the minimum variance estimator. It is therefore interesting to examine the difference between this estimator and the weighted and unweighted mean that can be used in practice more easily. We will also consider the plug-in estimator based on the multivariate analysis of variance estimates for $\boldsymbol{\Sigma}$ and $\mathbf{T}$, given by (14). Note that the plug-in estimator is a biased estimator, which motivates the use of the mean squared error to compare the mean estimators and not only the variance. To compare the performance of the toy estimator and the plug-in estimator with the performance of the weighted and unweighted mean, the simulations as described in Section 4.1 based on the same values of m and corresponding n_i's are used. The results of these Monte Carlo simulations are given in Table 1 and 2 and Figure 1.

An interesting observation from Table 1 and 2 is that the optimal generalized weighted mean has (approximately) the same mean squared errors as the unweighted mean in this simulation. This can be explained by the small value for the parameter $\boldsymbol{\Sigma}$ in comparison to the value for $\mathbf{T}$, see equation (15). Consequently, it follows that $\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma} \approx \mathbf{T}$. Hence,

$$\mathrm{Cov}\left(\hat{\boldsymbol{\mu}}_u, \hat{\boldsymbol{\mu}}_u\right) \approx \frac{1}{m^2}\sum_{i=1}^{m}\mathbf{T} = \frac{\mathbf{T}}{m} \quad \text{and therefore} \quad \mathrm{Var}(\hat{\boldsymbol{\mu}}_u) \approx \frac{tr(\mathbf{T})}{m}.$$

The weight matrices for the optimal generalized weighted mean are approximately equal to

$$\mathbf{W}_i \approx \left(\sum_{i=1}^{m}\mathbf{T}^{-1}\right)^{-1}\mathbf{T}^{-1} = \frac{1}{m}\mathbf{I}_k$$

so that the variance of the optimal generalized weighted mean is approximately

$$\mathrm{Var}\left(\hat{\boldsymbol{\mu}}_{\mathrm{opt}}\right) \approx \sum_{i=1}^{m} tr\left(\frac{1}{m}\mathbf{I}_k\mathbf{T}\frac{1}{m}\mathbf{I}_k\right) = \frac{tr(\mathbf{T})}{m}.$$

Thus, if the within-source variation is small relative to the between-source variation it follows that

$$\text{Var}\left(\hat{\boldsymbol{\mu}}_u\right) \approx \text{Var}\left(\hat{\boldsymbol{\mu}}_{\text{opt}}\right).$$

Hence, for such situations the unweighted mean is as good as the minimum variance estimator $\hat{\boldsymbol{\mu}}_{\text{opt}}$. Note that the plug-in estimator $\hat{\boldsymbol{\mu}}_{\text{plug}}$ also behaves similarly to the minimum variance estimator.

4.3 Artificial choice of parameters

For the covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{T}$ from equation (15), we have seen that the within-source variation $\boldsymbol{\Sigma}$ is very small so that $\mathbf{T}+n_i^{-1}\boldsymbol{\Sigma} \approx \mathbf{T}$ and therefore the unweighted mean is approximately as good as the minimum variance estimator $\hat{\boldsymbol{\mu}}_{\text{opt}}$. It is interesting to consider some situations where $\hat{\boldsymbol{\mu}}_{\text{opt}}$ outperforms both the weighted and unweighted mean estimator. To this end, the following artificial choice of parameters was made:

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.3 & 0.0 & 0.3 \\ 0.0 & 0.1 & -0.2 \\ 0.3 & -0.2 & 0.8 \end{bmatrix}, \qquad \mathbf{T} = \begin{bmatrix} 0.6 & 0.3 & 0.5 \\ 0.3 & 0.4 & 0.2 \\ 0.5 & 0.2 & 0.9 \end{bmatrix}. \tag{16}$$

Again a Monte Carlo simulation study is performed for $m = 1200$ groups, as was described in Section 4.1. The values of r and c for the simulated data set are equal to 0.632 and 3.49 respectively, so that the inequality $tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T})$ holds. The mean squared errors of both the mean estimates and the likelihood ratio values are given in Table 3.

	$\hat{\boldsymbol{\mu}}_w$	$\hat{\boldsymbol{\mu}}_u$	$\hat{\boldsymbol{\mu}}_{\text{opt}}$	$\hat{\boldsymbol{\mu}}_{\text{plug}}$
MSE $\hat{\boldsymbol{\mu}}$	$2.13 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$1.73 \cdot 10^{-3}$	$1.73 \cdot 10^{-3}$
MSE LR$(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2 \vert \hat{\boldsymbol{\mu}})$	$4.54 \cdot 10^{-3}$	$3.59 \cdot 10^{-3}$	$3.56 \cdot 10^{-3}$	$3.56 \cdot 10^{-3}$

Table 3: Mean squared errors of the estimated mean and likelihood ratio using the weighted mean $\hat{\boldsymbol{\mu}}_w$, the unweighted mean $\hat{\boldsymbol{\mu}}_u$, the toy estimator $\hat{\boldsymbol{\mu}}_{\text{opt}}$ and the plug-in estimator $\hat{\boldsymbol{\mu}}_{\text{plug}}$ for a Monte Carlo simulation ($M = 1000$) with parameters as given in equation (16) and $m = 1200$ groups. The true likelihood ratio is equal to 2.47.

Indeed, the optimal generalized weighted mean performs better than the other overall mean estimators, although the performance is comparable to that of the unweighted mean estimator and the plug-in mean estimator.

Another interesting situation is when the inequality $tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T})$ does not hold. Therefore, the parameter $\boldsymbol{\Sigma}$ is multiplied by 10 whereas the other parameters as well as the sampled n_i's remain unchanged. Again a Monte Carlo simulation study is performed for $m = 1200$ groups, but we now have $r = 6.32$ and $c = 3.49$ so that $tr(\boldsymbol{\Sigma}) > c \cdot tr(\mathbf{T})$. This means that the weighted mean should perform better than the unweighted mean. Note that the values of r and c do not influence $\hat{\boldsymbol{\mu}}_{\text{opt}}$ and that this is still the minimum variance unbiased estimator. The results of the simulation study can be found in Table 4.

	$\hat{\boldsymbol{\mu}}_w$	$\hat{\boldsymbol{\mu}}_u$	$\hat{\boldsymbol{\mu}}_{\text{opt}}$	$\hat{\boldsymbol{\mu}}_{\text{plug}}$
MSE $\hat{\boldsymbol{\mu}}$	$2.96 \cdot 10^{-3}$	$3.31 \cdot 10^{-3}$	$2.67 \cdot 10^{-3}$	$2.67 \cdot 10^{-3}$
MSE LR$(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2 \vert \hat{\boldsymbol{\mu}})$	$1.65 \cdot 10^{-2}$	$1.71 \cdot 10^{-2}$	$1.45 \cdot 10^{-2}$	$1.45 \cdot 10^{-2}$

Table 4: Mean squared errors of the estimated mean and likelihood ratio using the weighted mean $\hat{\boldsymbol{\mu}}_w$, the unweighted mean $\hat{\boldsymbol{\mu}}_u$, the toy estimator $\hat{\boldsymbol{\mu}}_{\text{opt}}$ and the plug-in estimator $\hat{\boldsymbol{\mu}}_{\text{plug}}$ for a Monte Carlo simulation ($M = 1000$) with parameters as given in equation (16), where $\boldsymbol{\Sigma}$ is multiplied by 10, and $m = 1200$ groups. The true likelihood ratio is equal to 4.58.

As expected, the weighted mean now performs better than the unweighted mean, but the optimal generalized weighted mean is still the best of all estimators. Again, the performance of the toy estimator and the plug-in estimator is similar.

5 Estimating the overall mean of XTC data

In this section, the different estimators will be applied to real XTC data to illustrate the impact of the choice of overall mean estimator. The XTC data come from the CHAMP (Collaborative Harmonization of Methods for Profiling of Amphetamine Type Stimulants) project. Instead of generating the control and recovered data $\mathbf{Y}_1$ and $\mathbf{Y}_2$ based on the parameters given in equation (15), it is also possible to apply the three mean estimators to real XTC trace evidence. Since the true mean $\boldsymbol{\mu}$ and the true likelihood ratio $\text{LR}(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2|\boldsymbol{\mu})$ are now unknown, we cannot say anything about mean squared errors. Therefore, this application is purely meant to indicate the difference in results when using the weighted, unweighted or optimal plug-in generalized weighted mean. The latter will again be based on the multivariate analysis of variance estimates as given in equation (14). In fact, these are the same estimates as used to obtain the parameters $\boldsymbol{\Sigma}$ and $\mathbf{T}$ in equation (15) from the real XTC data.

The control data $\mathbf{Y}_1$ now consists of 42 measurements of the diameter, thickness and weight of tablets from consignment C_1, and the recovered data $\mathbf{Y}_2$ consists of 5 measurements on tablets that also come from consignment C_1. This means that the prosecutor's hypothesis is true and likelihood ratio values larger than 1 are expected. It is assumed that the origin of consignment C_1 is unknown, i.e. it is not known which production process produced the tablets, so that indeed the described two-level model applies to this situation. The background data consists of 186 consignments with two or more tablet measurements where it is not known whether there are links between the consignments. For this data set, we have $c = 11.0$ and $r = tr(\boldsymbol{\Sigma})/tr(\mathbf{T}) = 0.0254$, so that the inequality $tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T})$ holds. The following estimates for the overall mean $\boldsymbol{\mu}$ are obtained:

$$\hat{\boldsymbol{\mu}}_w = \begin{bmatrix} 8.242 \\ 4.528 \\ 276.0 \end{bmatrix}, \qquad \hat{\boldsymbol{\mu}}_u = \begin{bmatrix} 8.240 \\ 4.211 \\ 260.0 \end{bmatrix}, \qquad \hat{\boldsymbol{\mu}}_{\text{plug}} = \begin{bmatrix} 8.240 \\ 4.212 \\ 260.1 \end{bmatrix}.$$

Using the same estimates from equations (14) for $\boldsymbol{\Sigma}$ and $\mathbf{T}$ and the likelihood ratio formula from equation (2), the likelihood ratio values can be calculated for each of the overall mean estimates:

$$\text{LR}(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2|\hat{\boldsymbol{\mu}}_w) = 1455, \qquad \text{LR}(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2|\hat{\boldsymbol{\mu}}_u) = 2073, \qquad \text{LR}(\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2|\hat{\boldsymbol{\mu}}_{\text{plug}}) = 2072.$$

This shows that there is a significant difference in likelihood ratio values when using $\hat{\boldsymbol{\mu}}_w$ instead of $\hat{\boldsymbol{\mu}}_u$ or $\hat{\boldsymbol{\mu}}_{\text{plug}}$. The analysis in the previous sections showed that, since $tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T})$, both $\hat{\boldsymbol{\mu}}_u$ and $\hat{\boldsymbol{\mu}}_{\text{plug}}$ outperform $\hat{\boldsymbol{\mu}}_w$. Hence, it would be strongly discouraged to use the weighted mean when reporting likelihood ratio values for this evidence set.

6 Conclusion

In this paper three estimators for the mean are presented, which can be used if the evidence is modelled as a two-level model using assumptions of multivariate normality: the weighted mean, the unweighted mean and a generalized weighted mean estimator. The choice of the estimator of the overall mean is important for the estimation of the between-source covariance matrix and for the calculation of the likelihood ratio. There is no consensus on which of these two estimators to use when the data are unbalanced. In this paper a relation is found which can be used to find the most efficient estimator and thus to decide whether the weighted or the unweighted mean should be used. The unweighted mean is preferred over the weighted mean if $tr(\boldsymbol{\Sigma}) < c \cdot tr(\mathbf{T})$, where the constant c depends on the number of groups in the background data and the number of

measurements in each group. It is argued that in many forensic comparison problems the within-source variation can be assumed to be smaller than the between-source variation. Moreover, it is proven that the value of c will never be smaller than one. Therefore, it is expected that in practice the unweighted mean will often be preferred over the weighted mean. Of course, there might also be contextual reasons to prefer one of the overall mean estimators over the other.

The weights of the generalized weighted mean are derived such that they minimize the variance of this estimator. These optimal weights lead to a toy estimator, because they depend on the unknown within- and between-source covariance matrices. If these parameters would be known, the derived toy estimator has smaller (or equal) variance than the weighted and the unweighted mean. Using the optimal weights with estimates for the within- and between-source covariance matrices leads to a plug-in estimator. When comparing the multivariate mean estimators in a simulation study where the unweighted mean should be preferred over the weighted mean, the unweighted mean and plug-in estimator perform similarly to the toy estimator which yields minimum variance. Using an artificial choice of parameters provides some examples where the toy estimator outperforms both the weighted and unweighted mean, regardless of the number of groups and number of measurements in the background data. Applying the weighted mean, the unweighted mean and the plug-in mean estimator to real data shows the impact that the choice of estimator has on the value of evidence.

A Appendix

A.1 The variance of $\hat{\boldsymbol{\mu}}_w$ and $\hat{\boldsymbol{\mu}}_u$

The covariance matrix of the weighted mean $\hat{\boldsymbol{\mu}}_w$ is found by setting $\mathbf{W}_i = \frac{n_i}{N}\mathbf{I}_k$ in equation (11) so that

$$\text{Cov}(\hat{\boldsymbol{\mu}}_w, \hat{\boldsymbol{\mu}}_w) = \sum_{i=1}^{m} \frac{n_i}{N}\mathbf{I}_k(\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})\frac{n_i}{N}\mathbf{I}_k = \frac{\mathbf{T}}{N^2}\sum_{i=1}^{m} n_i^2 + \frac{\boldsymbol{\Sigma}}{N}.$$

By linearity of the trace, we have

$$\text{Var}(\hat{\boldsymbol{\mu}}_w) = \frac{tr(\mathbf{T})}{N^2}\sum_{i=1}^{m} n_i^2 + \frac{tr(\boldsymbol{\Sigma})}{N}.$$

Similarly, the covariance matrix of the unweighted mean $\hat{\boldsymbol{\mu}}_u$ can be found by setting $\mathbf{W}_i = \frac{1}{m}\mathbf{I}_k$ in equation (11) so that

$$\text{Cov}(\hat{\boldsymbol{\mu}}_u, \hat{\boldsymbol{\mu}}_u) = \sum_{i=1}^{m} \frac{1}{m}\mathbf{I}_k(\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})\frac{1}{m}\mathbf{I}_k = \frac{\mathbf{T}}{m} + \frac{\boldsymbol{\Sigma}}{m^2}\sum_{i=1}^{m} \frac{1}{n_i}$$

and linearity of the trace gives

$$\text{Var}(\hat{\boldsymbol{\mu}}_u) = \frac{tr(\mathbf{T})}{m} + \frac{tr(\boldsymbol{\Sigma})}{m^2}\sum_{i=1}^{m} \frac{1}{n_i}.$$

A.2 The efficiency of $\hat{\boldsymbol{\mu}}_u$ relative to $\hat{\boldsymbol{\mu}}_w$

To find the relation of the efficiency as given in expression (9), Jensen's inequality can be used. Consider the random variable U, uniformly distributed on $n_1, \ldots, n_m$, ordered integers ≥ 1. By Jensen's inequality it follows that

$$\frac{1}{m}\sum_{i=1}^{m} \frac{1}{n_i} = \text{E}\left[\frac{1}{U}\right] \geq \frac{1}{\text{E}[U]} = \frac{1}{\frac{1}{m}\sum_{i=1}^{m} n_i} = \frac{m}{N}. \tag{17}$$

Here is used that the function $\phi(x) = \frac{1}{x}$ is convex for $x > 0$, which is sufficient since only positive values are considered. Applying Jensen's inequality to the function $\phi(x) = x^2$ it follows that

$$\frac{1}{m}\sum_{i=1}^{m} n_i^2 = \mathrm{E}\left[U^2\right] \geq (\mathrm{E}[U])^2 = \left(\frac{1}{m}\sum_{i=1}^{m} n_i\right)^2 = \frac{N^2}{m^2}. \tag{18}$$

Moreover, from inequality (18) it follows that

$$\frac{tr(\mathbf{T})}{N^2}\sum_{i=1}^{m} n_i^2 \geq \frac{tr(\mathbf{T})}{N^2}\frac{mN^2}{m^2} = \frac{tr(\mathbf{T})}{m},$$

which refers to the first terms in the numerator and denominator of equation (6). On the other hand, from inequality (17) it follows that

$$\frac{tr(\mathbf{\Sigma})}{m^2}\sum_{i=1}^{m} \frac{1}{n_i} \geq \frac{tr(\mathbf{\Sigma})}{N},$$

which refers to the second terms in the denominator and numerator of equation (6).

A.3 Proof of Lemma 3.1

Multiplying both the numerator and the denominator by $\frac{1}{m^3}$ and using $N = \sum_{i=1}^{m} n_i$, the expression for c can be re-written to

$$c = \frac{\frac{1}{m}\sum_{i=1}^{m} n_i^2 - \left(\frac{1}{m}\sum_{i=1}^{m} n_i\right)^2}{\left(\frac{1}{m}\sum_{i=1}^{m} n_i\right)^2 \left(\frac{1}{m}\sum_{i=1}^{m} \frac{1}{n_i}\right) - \frac{1}{m}\sum_{i=1}^{m} n_i}.$$

To simplify notation a bit, consider the random variable U as defined in Appendix A.2. Then we can write

$$c^{-1} = \frac{(\mathrm{E}[U])^2\,\mathrm{E}[U^{-1}] - \mathrm{E}[U]}{\mathrm{Var}(U)}.$$

Consider the convex function ϕ on $[1, n_m]$ defined by

$$\phi(y) = y^{-1}.$$

Since ϕ is a convex function, the tangent lines to ϕ are below the graph of ϕ. The idea of the proof is to find a parabola that can be added to the tangent lines so that it will always be above the graph of ϕ, see Figure 2. It follows that for fixed $u \in [1, n_m]$ we have for any $y \in [1, n_m]$

$$\phi(y) \leq \phi(u) + \phi'(u)(y-u) + \frac{(y-u)^2}{u^2}. \tag{19}$$

Indeed,

$$\frac{1}{y} \leq \frac{1}{u} - \frac{1}{u^2}(y-u) + \frac{1}{u^2}(y-u)^2,$$

which can be re-written to

$$\frac{(y-1)(u-y)^2}{u^2 y} \geq 0$$

and holds as long as $u \geq 1$ and $y \geq 1$.

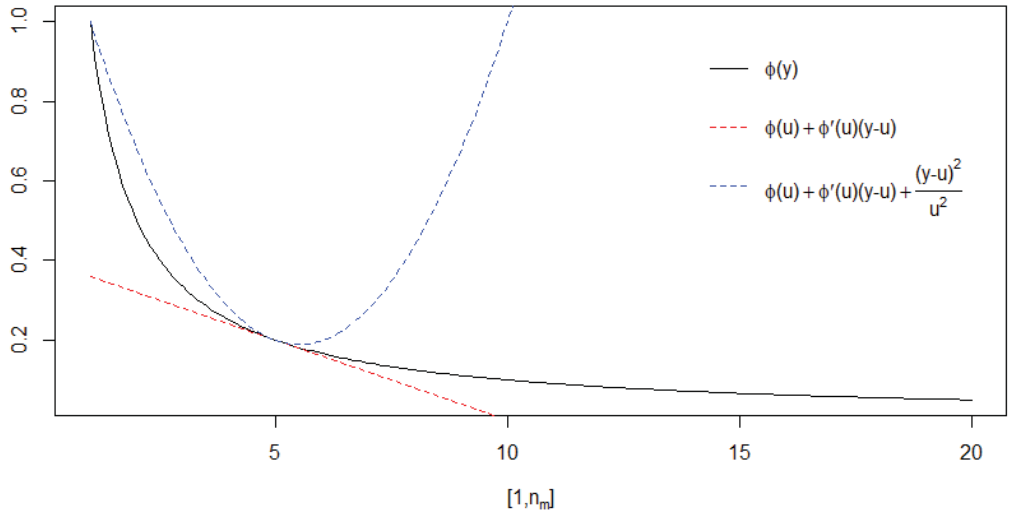

Figure 2: Illustration of equation (19) for $u = 5$.

Choosing $u = \mathrm{E}[U] \geq 1$ and substituting the random variable $U \geq 1$ for y results in

$$\frac{1}{U} \leq \frac{1}{\mathrm{E}[U]} - \frac{1}{(\mathrm{E}[U])^2}(U - \mathrm{E}[U]) + \frac{1}{(\mathrm{E}[U])^2}(U - \mathrm{E}[U])^2.$$

Now taking expectations, we get

$$\mathrm{E}[U^{-1}] \leq \frac{1}{\mathrm{E}[U]} + \frac{1}{(\mathrm{E}[U])^2}\,\mathrm{Var}(U).$$

Hence,

$$(\mathrm{E}[U])^2\,\mathrm{E}[U]^{-1} - \mathrm{E}[U] \leq \mathrm{Var}(U),$$

which implies that $c^{-1} \leq 1$, i.e. $c \geq 1$.

A.4 Proof of Lemma 3.2

To minimize $\mathrm{Var}(\hat{\boldsymbol{\mu}})$ subject to the constraint $\mathbf{W}_1 + \cdots + \mathbf{W}_m = \mathbf{I}_k$ a k^2-dimensional Lagrange multiplier $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \ldots, \lambda_{kk})$ is introduced such that the Lagrange function is equal to:

$$\mathcal{L}_{\boldsymbol{\lambda}}(\mathbf{W}_1, \ldots, \mathbf{W}_m, \lambda_{11}, \lambda_{12}, \ldots, \lambda_{kk}) = f(\mathbf{W}_1, \ldots, \mathbf{W}_m) - \sum_{j=1}^{k}\sum_{l=1}^{k} \lambda_{jl} g_{jl}(\mathbf{W}_1, \ldots, \mathbf{W}_m),$$

where

$$f(\mathbf{W}_1, \ldots, \mathbf{W}_m) = \sum_{i=1}^{m} tr\left(\mathbf{W}_i(\mathbf{T} + n_i^{-1}\boldsymbol{\Sigma})\mathbf{W}_i^T\right)$$

and

$$g_{jl}(\mathbf{W}_1, \ldots, \mathbf{W}_m) = \left[\sum_{i=1}^{m} \mathbf{W}_i - \mathbf{I}_k\right]_{jl}.$$

I.e. $g_{jl}(\mathbf{W}_1, \ldots, \mathbf{W}_m)$ is equal to the matrix element with index jl. Let

$$\frac{\partial}{\partial \mathbf{W}_i} = \begin{bmatrix} \frac{\partial}{\partial w_{i,11}} & \frac{\partial}{\partial w_{i,12}} & \cdots & \frac{\partial}{\partial w_{i,1k}} \\ \frac{\partial}{\partial w_{i,21}} & \frac{\partial}{\partial w_{i,22}} & \cdots & \frac{\partial}{\partial w_{i,2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial w_{i,k1}} & \frac{\partial}{\partial w_{i,k2}} & \cdots & \frac{\partial}{\partial w_{i,kk}} \end{bmatrix}$$

denote the derivative with respect to the matrix $\mathbf{W}_i$. Then we have

$$\frac{\partial}{\partial \mathbf{W}_i}\left(\sum_{j=1}^{m} tr\left(\mathbf{W}_j(\mathbf{T} + n_j^{-1}\mathbf{\Sigma})\mathbf{W}_j^T\right)\right) = \frac{\partial}{\partial \mathbf{W}_i}\, tr\left(\mathbf{W}_i(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})\mathbf{W}_i^T\right) = 2\mathbf{W}_i(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})$$

since $(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})$ is symmetric and $\dfrac{\partial\, tr(\mathbf{X}\mathbf{A}\mathbf{X}^T)}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{A} + \mathbf{A}^T)$ [7]. Clearly,

$$\frac{\partial}{\partial w_{i,jl}} g_{jl}(\mathbf{W}_1, \ldots, \mathbf{W}_m) = 1$$

and zero for all other indices. Therefore, it follows that

$$\frac{\partial \mathcal{L}_{\boldsymbol{\lambda}}}{\partial \mathbf{W}_i} = 2\mathbf{W}_i(\mathbf{T} + n_i^{-1}\mathbf{\Sigma}) - \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k1} & \lambda_{k2} & \cdots & \lambda_{kk} \end{bmatrix} := 2\mathbf{W}_i(\mathbf{T} + n_i^{-1}\mathbf{\Sigma}) - \mathbf{\Lambda}$$

and the Lagrange function will be minimized over $\mathbb{R}^{k\times k}$. Setting the derivative equal to the $k \times k$ zero matrix results in

$$\mathbf{W}_i = \frac{1}{2}\mathbf{\Lambda}(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})^{-1}, \qquad 1 \le i \le m.$$

Now using the constraint $\sum_{i=1}^{m} \mathbf{W}_i = \mathbf{I}_k$ gives

$$\sum_{i=1}^{m} \frac{1}{2}\mathbf{\Lambda}(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})^{-1} = \mathbf{I}_k.$$

Hence,

$$\frac{1}{2}\mathbf{\Lambda} = \left(\sum_{i=1}^{m}(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})^{-1}\right)^{-1}.$$

Thus,

$$\mathbf{W}_i = \left(\sum_{j=1}^{m}(\mathbf{T} + n_j^{-1}\mathbf{\Sigma})^{-1}\right)^{-1}(\mathbf{T} + n_i^{-1}\mathbf{\Sigma})^{-1}, \qquad 1 \le i \le m$$

which proves the lemma.

References

[1] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Appl. Stat. 53 Part 1 (2004) 109-122.

[2] C.G.G. Aitken, F. Taroni, Statistics and the evaluation of evidence for forensic scientists, second ed., John Wiley & Sons, United Kingdom, 2004, pp. 319-381.

[3] I. Alberink, A. Bolck, S. Menges, Posterior likelihood ratio for evaluation of forensic trace evidence given a two-level model on the data, Appl. Stat. 40 (2013) 2579-2600.

[4] A. Bolck, C. Weyermann, L. Dujourdy, P. Esseiva, J. van den Berg, Different likelihood ratio approaches to evaluate the strength of evidence of XTC tablet comparisons, Forensic Science International 191 (2009) 42-51.

[5] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence, likelihood ratio for autorship, Appl. Stat. 57 (2008) 329-343.

[6] D.V. Lindley, A problem in forensic science, Biometrika 64 (1977) No. 2 207-213.

[7] K.B. Petersen, M.S. Pedersen, The Matrix Cookbook, 2012. URL `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

[8] D.M. Ommen, Approximate statistical solutions to the forensic identification of source problem, 2017, South Dakota State University.

[9] D.M. Ommen, C.P. Saunders, Building a unified statistical framework for the forensic identification of source problems, Law, Probability and Risk (2018).

[10] A. Rice, Mathematical statistics and data analysis, first ed., Brook/Cole, Berkeley, 2007, pp. 136,227-238.

[11] R. Royall, Statistical Evidence: A likelihood paradigm, first ed., Chapman & Hall, London, 1997.

[12] H. Sahai, M. Ojeda, Analysis of variance for random models, Unbalanced Data, Theory, Methods, Applications and Data Analysis, Volume II, first ed., Birkhäuser, Boston, 2005, pp. 102-103.

[13] S.R. Searle, G. Casella, C.E. McCulloch, Variance components, first ed., John Wiley & Sons, New York, 1992, pp. 44-90.

[14] F. Taroni, S. Bozza, A. Biederman, P. Garbolino, C. Aitken, Data analysis in forensic science, A Bayesian decision perspective, first ed., John Wiley & Sons, United Kingdom, 2010, pp. 137-185.

[15] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Statistical analysis in forensic science, Evidential value of multivariate physicochemical data, first ed., John Wiley & Sons, United Kingdom, 2014, pp. 107-110.