Ann. Inst. Statist. Math.
Vol. 54, No. 4, 827–839 (2002)

EFFICIENT NON-ITERATIVE AND NONPARAMETRIC ESTIMATION OF HETEROGENEITY VARIANCE FOR THE STANDARDIZED MORTALITY RATIO

DANKMAR BÖHNING[1], UWE MALZAHN[1], JESUS SAROL, JR.[2], SASIVIMOL RATTANASIRI[3] AND ANNIBALE BIGGERI[4]

[1]*Department of Epidemiology, Free University Berlin, Haus 562, Fabeckstr. 60-62, 14195 Berlin, Germany,* e-mail: boehning@zedat.fu-berlin.de
[2]*Department of Epidemiology and Biostatistics, College of Public Health, University of the Philippines Manila, Manila, Philippines,* e-mail: jsarol@nwave.net
[3]*Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand,* e-mail: r_sasivimol@hotmail.com
[4]*Department of Statistics, University of Florence, 50134 Florence, Italy,* e-mail: abiggeri@stat.ds.unifi.it

(Received December 5, 2000; revised August 6, 2001)

Abstract. In this paper the situation of extra population heterogeneity in the standardized mortality ratio is discussed from the point-of-view of an analysis of variance. First, some simple non-iterative ways are provided to estimate the variance of the heterogeneity distribution without estimating the heterogeneity distribution itself. Next, a wider class of linear unbiased estimators is introduced and their properties investigated. Consistency is shown for a wide sub-class of estimators charactererized by the fact that the associated linear weights are within some positive, finite bounds. Furthermore, it is shown that an efficient estimator is often provided when the weights are proportional to the expected counts.

Key words and phrases: Population heterogeneity, random effects model, moment estimator, variance separation, standardized mortality ratio.

1. Introduction

In a variety of biometric applications the situation of extra-population heterogeneity occurs. This is particularly the case if a good reason exists to model the variable of interest Y through a density of parametric form $p(y \mid \theta)$ with a scalar parameter θ. For a given subpopulation, the density $p(y \mid \theta)$ might be most suitable, but the value of θ cannot cover the whole population of interest. In such situations we speak of extra heterogeneity, which might be caused by unobserved covariates or clustered observations, such as herd clustering when estimating animal infection rates. An introductory discussion can be found in Aitkin *et al.* ((1990), p. 213) and the references given there; see also the review of Pendergast *et al.* ((1996), p. 106). A discussion on extra-binomial variation (i.e. extra-population heterogeneity if $p(y \mid \theta)$ is the binomial) can be found in Williams (1982) and Collet ((1991), p. 192). In this paper, it is understood that extra-population heterogeneity, or in short, population heterogeneity, refers to a situation when the parameter of interest, θ, varies in the population and sampling has not taken this into account (e.g. it has not been observed from which subpopulation (defined by the values of θ) the datum is coming from). As will be clear from equation (1.1) below, inference is affected by the occurrence of extra-population heterogeneity. For example, variances of estimators of interest are often greatly increased, leading to wider confidence intervals as compared to conventional ones. To adjust these variances the estimation of the variance of the distribution associated with the extra-heterogeneity is required. The main objective of this paper is to present a moment estimator for the heterogeneity variance in a simple manner. To be more precise, if θ varies in itself with distribution G and associated density $g(\theta)$, the (unconditional) marginal density of Y can be given as $f(y) = \int_\Theta p(y \mid \theta)g(\theta)d\theta$. Of interest is the separation of the (unconditional) variance of Y (e.g. variance of Y with respect to $f(y)$) into two terms:

$$\text{(1.1)} \qquad \operatorname{Var}(Y) = \int_\Theta \operatorname{Var}(Y \mid \theta)g(\theta)d\theta + \int_\Theta (\mu(\theta) - \mu_Y)^2 g(\theta)d\theta$$

where $\mu(\theta)$ is the $E(Y \mid \theta)$ and $\mu_Y = \int y f(y)dy$ is the marginal mean of Y. Note that $\mu_Y = E_G(\mu(\theta))$. Note that we can also write (1.1) briefly as

$$\operatorname{Var}(Y) = E_G(\sigma^2(\theta)) + \operatorname{Var}_G(\mu(\theta)).$$

In the sequel we will also denote $\operatorname{Var}_G(\mu(\theta))$ by τ_Y^2. Thus, in such instances, it can be said that (1.1) is a partitioning of the variance due to the variation in the subpopulation with parameter value θ (and then averaged over θ) and due to the variance in the heterogeneity distribution G of θ. Also, (1.1) can be taken as an analysis-of-variance partition with a latent factor with distribution G. We have to distinguish carefully between *three* distributional schemes when computing moments. For example, $\operatorname{Var}(Y)$ refers to the unconditional or marginal variance and is computed using the marginal density $f(y)$, $\operatorname{Var}(Y \mid \theta)$ is the *conditional* variance and is computed using the conditional density $p(y \mid \theta)$, and $\operatorname{Var}_G(\mu(\theta))$ refers to the distribution G of θ. The intention is to find an estimate of τ_Y^2 without implying knowledge or estimating the latent heterogeneity distribution G. The idea is very simple: we write (1.1) as

$$\text{(1.2)} \qquad \operatorname{Var}_G(\mu(\theta)) = \tau_Y^2 = \operatorname{Var}(Y) - E_G(\sigma^2(\theta))$$

and replace $\operatorname{Var}(Y)$ and $E_G(\sigma^2(\theta))$ on the right hand side of (1.2) with their respective sample estimates and obtain an estimate for τ_Y^2. In the succeeding text, we will use μ as the mean of θ and τ^2 for its variance.

Example (Poisson). Let $Y_1, Y_2, \ldots, Y_N$ be a random sample of Poisson counts, e.g. $p(y \mid \theta) = \exp(-\theta)\theta^y/y!$. Then, $\sigma^2(\theta) = \theta$, $E_G(\sigma^2(\theta)) = E_G(\theta) = \mu = E(Y)$ and $\tau_Y^2 = \tau^2$. Note that $\operatorname{Var}(Y)$ can simply be estimated by $S^2 = \frac{1}{N-1}\sum_{i=1}^N (Y_i - \bar{Y})^2$ and μ by $\bar{Y}$. Therefore, according to (1.2), an estimator of τ^2 is provided as $\hat{\tau^2} = S^2 - \bar{Y}$. This quantity has also been referred to as a measure of Poisson overdispersion (Böhning 1994). Note, that $E(\hat{\tau^2}) = \tau^2$.

Example (Binomial). Let $Y_1, Y_2, \ldots, Y_N$ be a random sample of Binomial counts, e.g. $p(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{(n-y)}$. Then, $\mu(\theta) = n\theta$ and $\sigma^2(\theta) = n\theta(1-\theta)$. Also, $\tau_Y^2 = n^2\tau^2$. It follows that $E_G(n\theta) = n\mu$, $E_G(\sigma^2(\theta)) = nE_G(\theta - \theta^2) = n(\mu - E_G(\theta^2)) = n(\mu - \tau^2 - \mu^2)$. Since $\operatorname{Var}(Y_i) = E_G(\sigma^2(\theta)) + \tau_Y^2 = n\mu(1-\mu) + n(n-1)\tau^2$, we find $\tau^2 = \frac{1}{n(n-1)}[\text{Var}(Y_i) - n\mu(1-\mu)]$, for $i = 1, \ldots, N$. We can use the estimator, $\hat{\tau^2} = \frac{S^2}{n(n-1)} - [\frac{\bar{Y}}{n}(1 - \frac{\bar{Y}}{n}]/(n-1)$, with $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$. This estimator has a bias equal to $\text{Var}(Y_i)/[n^2(n-1)N]$ which is practically negligible even for moderate values of n. For example, if $n = 10$ and $N = 10$, then the bias of $\hat{\tau^2}$ is equal to 1/9000 of the variance of Y_i.

The idea to construct a simple moment estimator using equation (1.2) can be found in various instances in the literature including Marshall (1991) and Martuzzi and Elliot (1996). The latter considered the case that $p(y \mid \theta)$ is the binomial. However, the way this moment estimator is constructed is not unique. In this paper, we try to develop a more general framework for these kinds of estimators.

In the next section, we will consider a generalization of this idea to the standardized mortality ratio. In Section 3, we will discuss a more general class of linear unbiased estimators of the heterogeneity variance and provide a closed form expression for its variance. This enables us to provide a closed form expression for the efficient estimator. In Section 4, we will provide simple conditions for consistency. Section 5 considers estimating simultaneously the mean and variance of the heterogneity distribution. Section 6 ends the paper with a discussion of the results.

2. The standardized mortality ratio

We consider a special but important case. Let $Y_1, Y_2, \ldots, Y_N$ be a sample of counts which can be thought of as a sequence of mortality or morbidity cases. For each Y_i there exists a connected non-random number e_i, for $i = 1, \ldots, N$, which is interpreted as an expected number of counts and usually calculated on the basis of an external reference population. With the help of these numbers one can define the standardized mortality ratio as $SMR_i = Y_i/e_i$ and its expected value $E(SMR_i \mid \theta_i) = \theta_i$, for $i = 1, \ldots, N$. Frequently, this sample is coming from N geographic regions or areas. Therefore, this situation is closely related to the so-called field of *disease mapping*. For an introduction to this field see Böhning (2000) or Lawson *et al.* (1999).

Furthermore, conditionally on the value of θ, a Poisson distribution is assumed for $Y \mid \theta$: $p(y_i \mid \theta, e_i) = \exp(-\theta e_i)(\theta e_i)^{y_i}/y_i!$. For this case, the partition of variance (1.1) takes the form

$$\begin{aligned} \text{Var}(Y_i) &= E_G(\sigma_i^2(\theta)) + \text{Var}_G(\mu_i(\theta)) = e_i E_G(\theta) + e_i^2 \, \text{Var}_G(\theta) \\ &= e_i\mu + e_i^2\tau^2. \end{aligned} \tag{2.1}$$

At this point it is important to understand the consequences of the occurrence of heterogeneity. Suppose μ is estimated using the conventional estimator $\hat{\mu} = \frac{\sum_i Y_i}{\sum_i e_i}$. Then, we have that $\text{Var}(\hat{\mu}) = \mu\frac{1}{\sum_i e_i} + \tau^2\frac{\sum_i e_i^2}{(\sum_i e_i)^2}$, so that, depending on the value of τ^2, its variance might be largely increased. Note also that conventional confidence intervals use the variance formula $\text{Var}(\hat{\mu}) = \mu\frac{1}{\sum_i e_i}$, which might be too small if heterogeneity is present.

We write (2.1) as $E(Y_i - e_i\mu)^2 = e_i\mu + e_i^2\tau^2$ which draws attention to the variate $W_i = \frac{(Y_i - e_i\mu)^2 - e_i\mu}{e_i^2}$. Since $\text{Var}(Y_i) = E(Y_i - e_i\mu)^2$ we note that it follows from (2.1)

$$E(W_i) = \tau^2. \tag{2.2}$$

First, to estimate τ^2, we can replace $\text{Var}(Y_i)$ by its 'estimate' $(Y_i - e_i\mu)^2$ and solve for τ^2 and then average over i:

$$\hat{\tau}_1^2 = \frac{1}{N}\left[\sum_{i=1}^{N}(Y_i - e_i\mu)^2/e_i^2 - \mu\sum_{i=1}^{N}\frac{1}{e_i}\right]. \tag{2.3}$$

Second, in (2.1), we can divide first by e_i and then average over i and solve for τ^2:

$$\hat{\tau}_2^2 = \frac{\sum_{i=1}^{N}(Y_i - e_i\mu)^2/e_i - \mu N}{\sum_{i=1}^{N} e_i}. \tag{2.4}$$

Third, we can also first average over i in (2.1), and then solve for τ^2:

$$\hat{\tau}_3^2 = \frac{\sum_{i=1}^{N}(Y_i - e_i\mu)^2 - \mu\sum_{i=1}^{N} e_i}{\sum_{i=1}^{N} e_i^2}. \tag{2.5}$$

Note that all three estimators are identical if the e_i's are all equal (e.g. if $e_i = e_j$ for all $i, j = 1, \ldots, N$). We note in passing that all three estimators are unbiased. In fact, they are special cases of a more general class of *linear unbiased estimators* of τ^2:

$$T(W, \alpha) = \frac{\sum_{i=1}^{N}\alpha_i W_i}{\sum_{i=1}^{N}\alpha_i} \tag{2.6}$$

for any non-random, non-negative numbers $\alpha_1, \alpha_2, \ldots, \alpha_N$. It is easy to verify that for $\alpha_i = 1/N$ the estimator $T(W, \alpha) = \hat{\tau}_1^2$, for $\alpha_i = e_i$ the estimator $T(W, \alpha) = \hat{\tau}_2^2$, and for $\alpha_i = e_i^2$ the estimator $T(W, \alpha) = \hat{\tau}_3^2$ is provided. The estimator $\hat{\tau}_1^2$ associated with $\alpha_i = 1/N$ is mentioned in Böhning (2000). The estimator $\hat{\tau}_2^2$ associated with $\alpha_i = e_i$ is suggested by Marshall (1991).

The estimator $T(W, \alpha)$ considered so far requires the knowledge of the overall-mean μ. This assumption is satisfied, if the SMR_is are *indirectly standardized* implying that $\sum_i Y_i / \sum_i e_i = 1$.

2.1 *Example* 1: *Hepatitis B in Berlin*

To illustrate the estimators, we consider two examples. Table 1 gives the observed and expected Hepatitis B cases in the 23 city regions of Berlin for the year 1995. Here, we find that $\sum_i Y_i / \sum_i e_i = 1.019$. A conventional χ^2-test for homogeneity is given by $\chi^2 = \sum_i (Y_i - \mu e_i)^2/(\mu e_i)$. If μ is replaced with $\hat{\mu} = \sum_i Y_i / \sum_i e_i = 1.019$, we will get $\chi^2 = 193.52$, which clearly indicates heterogeneity. For this illustration, assuming that μ is fixed, the following values for $\hat{\tau}_j^2$ can be achieved: 0.5205($j = 1$), 0.4810($j = 2$), 0.4226($j = 3$). This indicates rather high heterogeneity since $\widehat{\text{Var}(SMR)} = \frac{1}{N-1}\sum_i (SMR_i - \overline{SMR})^2 = 0.6234$. The situation is illustrated in Fig. 1 (using $\hat{\tau}_1^2$ to construct the confidence interval adjusting for heterogeneity). Note that using the "right" estimate of variance leads to an increased length in confidence interval for μ using $\hat{\mu} \pm 1.96\sqrt{\text{Var}(\hat{\mu})}$ for the construction of a 95%-confidence interval where $\hat{\mu}$ corresponds to the pooled estimator.

Table 1. Observed and expected Hepatitis B cases in the 23 city regions of Berlin (1995).

Area i	Y_i	e_i	Area i	Y_i	e_i
1	29	10.7121	13	15	8.3969
2	26	17.9929	143	11	15.6438
3	54	18.1699	15	11	11.8289
4	30	19.2110	16	2	9.9513
5	16	21.9611	17	2	10.8313
6	15	14.6268	18	9	18.3404
7	6	9.6220	19	2	5.1758
8	35	17.2671	20	3	10.9543
9	17	18.8230	21	11	20.0121
10	7	18.2705	22	5	13.8389
11	43	32.1823	23	2	12.7996
12	17	24.5929	-	-	-

Source: Berlin Census Bureau

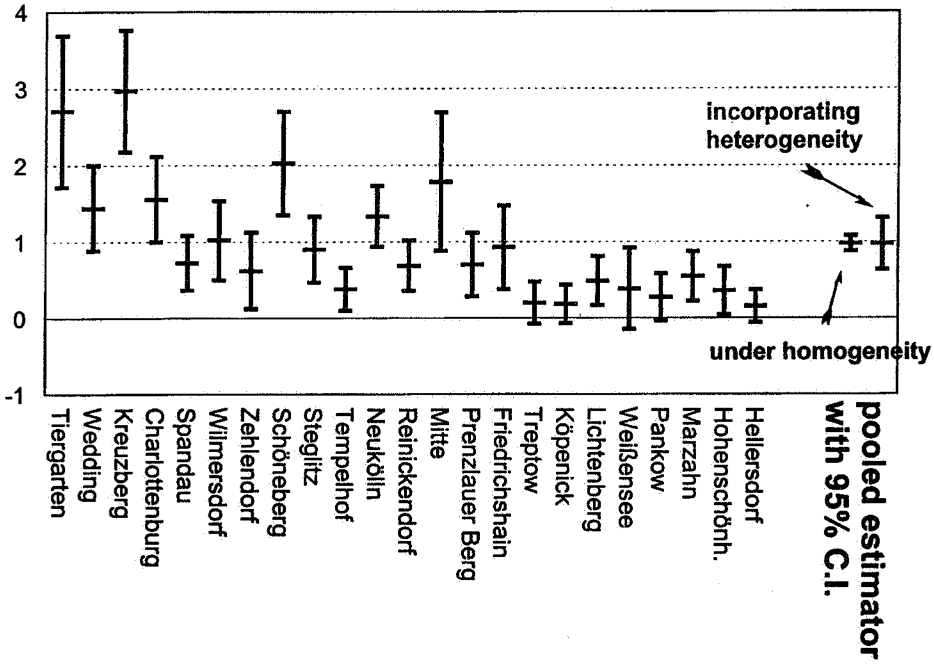

Fig. 1. SMR estimates of Hepatitis B in 23 Berlin city areas with pointwise 95%-confidence intervals.

2.2 *Example 2: Perinatal mortality in the North-West Thames Health Region*

As another realistic data set the small area data of Martuzzi and Hills (1995) on perinatal mortality in the North-West Thames Health Region in England based on the 5-year period 1986-1990 is considered. The region consists of 515 small areas. In this case, $\sum_i Y_i = \sum_i e_i = 2051$. It was found that $\hat{\tau}_1^2 = -0.0272790$ which is truncated to

0, and $\hat{\tau}_2^2 = 0.0167823$ as well as $\hat{\tau}_3^2 = 0.0369576$. There is small heterogeneity present in the data indicated by the ratio $\frac{\hat{\tau}_i^2}{\widehat{\text{Var}(SMR)}}$, where $\widehat{\text{Var}(SMR)} = \frac{1}{N-1}\sum_i (SMR_i - \overline{SMR})^2$, which takes on the values $0, \frac{0.0168}{0.6058}, \frac{0.0370}{0.6058} = 0, 0.0277, 0.0611$ for the 3 estimators, respectively.

3. Efficiency

When investigating the efficiency of the family of estimators $T(W,\alpha)$, we have to consider its variance:

$$\text{Var}(T(W,\alpha)) = \frac{\sum_i \alpha_i^2 \text{Var}(W_i)}{(\sum_i \alpha_i)^2} \tag{3.1}$$

which is completely specified, if $\text{Var}(W_i)$ is known. It is well-known that the efficient estimator (i.e. the one with minimum variance in the family $T(W,\alpha)$) chooses α_i proportional to $\frac{1}{\text{Var}(W_i)}$. Consequently, our interest concentrates on $\text{Var}(W_i)$. We have the following result.

Lemma 3.1. *Let G be any distribution with finite moments to the power of four. Then:*

$$\begin{aligned}\text{Var}(W_i) &= \mu e_i^{-3} + (2\mu^2 + 7\tau^2)e_i^{-2} + 2(3\mu^{(3)} - 7\mu\tau^2 - 3\mu^3)e_i^{-1} + 3\mu^4 \\ &\quad + \mu^{(4)} - \tau^4 + 6\mu^2\tau^2 - 4\mu\mu^{(3)}\end{aligned}$$

with $\mu^{(l)} = E_G(\theta^l)$ for $l = 3, 4$.

Proof. Note that $W_i = e_i^{-2}(Y_i - e_i\mu)^2 - e_i^{-1}\mu$, where μ is non- random and known. Consequently we have

$$\begin{aligned}\text{Var}(W_i) &= e_i^{-4}\text{Var}\{(Y_i - e_i\mu)^2\} \\ &= e_i^{-4}[E\{(Y_i - e_i\mu)^4\} - (E\{(Y_i - e_i\mu)^2\})^2].\end{aligned} \tag{3.2}$$

Note that for fixed θ_i the random variable Y_i is distributed according to the Poisson distribution with parameter $\theta_i e_i : Y_i \mid \theta_i \sim Po(\theta_i e_i)$. The moments up to the order of four for a Poisson distributed variable Y are needed here to use (3.2). These can be easily derived by the factorial moments. In Haight ((1967), p. 5–6) the moments are given up to the order of ten. In our application it follows

$$E(Y_i \mid \theta_i) = e_i\theta_i, \quad E(Y_i^2 \mid \theta_i) = e_i\theta_i + e_i^2\theta_i^2, \quad E(Y_i^3 \mid \theta_i) = e_i\theta_i + 3e_i^2\theta_i^2 + e_i^3\theta_i^3,$$
$$E(Y_i^4 \mid \theta_i) = e_i\theta_i + 7e_i^2\theta_i^2 + 6e_i^3\theta_i^3 + e_i^4\theta_i^4.$$

Furthermore, for each i the expected value of the SMR, θ_i, is to be interpreted as a realisation of the heterogeneity distribution $G : \theta_i \sim G$. Therefore, we have $E(Y_i^l) = E_G\{E\{Y_i^l \mid \theta_i\}\}, l = 1, 2, 3, 4$. From this fact, the moments of Y_i up to the power of four follow using the notation $\mu^{(l)} = E_G(\theta^l)$, $\mu = \mu^{(1)}$, $\tau^2 = \text{Var}_G(\theta) = \mu^{(2)} - \mu^2$:

$$\begin{aligned}&E(Y_i) = e_i\mu, \quad E(Y_i^2) = e_i\mu + e_i^2(\mu^2 + \tau^2), \\ &E(Y_i^3) = e_i\mu + 3e_i^2(\mu^2 + \tau^2) + e_i^3\mu^{(3)}, \\ &E(Y_i^4) = e_i\mu + 7e_i^2(\mu^2 + \tau^2) + 6e_i^3\mu^{(3)} + e_i^4\mu^{(4)}.\end{aligned}$$

Consequently, we have:

$$E\{(Y_i - e_i\mu)^2\} = e_i\mu + e_i^2\tau^2, \tag{3.3}$$

$$\begin{aligned} E\{(Y_i - e_i\mu)^4\} &= E(Y_i^4 - 4e_i\mu Y_i^3 + 6e_i^2\mu^2 Y_i^2 - 4e_i^3\mu^3 Y_i + e_i^4\mu^4) \\ &= e_i\mu + e_i^2(3\mu^2 + 7\tau^2) + 6e_i^3(\mu^{(3)} - 2\mu\tau^2 - \mu^3) \\ &\quad + e_i^4(\mu^{(4)} - 4\mu\mu^{(3)} + 6\mu^2\tau^2 + 3\mu^4). \end{aligned} \tag{3.4}$$

From (3.2), (3.3) and (3.4), we obtain the expression for $\operatorname{Var}(W_i)$ stated above. This ends the proof.

As a consequence from the expression for the variance of W_i derived above, it follows, that for large e_i, $\operatorname{Var}(W_i)$ behaves like a linear function in e_i^{-1}. To see this, note that

$$\frac{\partial}{\partial e_i^{-1}} \operatorname{Var}(W_i) = 3\mu e_i^{-2} + 2(2\mu^2 + 7\tau^2)e_i^{-1} + 6\mu^{(3)} - 14\mu\tau^2 - 6\mu^3.$$

Consequently, we have

$$\frac{\partial}{\partial e_i^{-1}} \operatorname{Var}(W_i) \to 2(3\mu^{(3)} - 7\mu\tau^2 - 3\mu^3) \quad \text{for} \quad e_i^{-1} \to 0.$$

This fact implies that, if we consider any fixed set of moments $(\mu, \tau^2, \mu^{(3)}, \mu^{(4)})$ and $\operatorname{Var}(W_i)$ as a function in e_i, then $\operatorname{Var}(W_i)$ increases approximately linearly with e_i^{-1} for large e_i. This result can be summarized in the following corollary.

COROLLARY 3.1.

$$\operatorname{Var}(W_i) \approx e_i^{-1} \quad \textit{for large} \quad e_i.$$

A further demonstration of this efficiency result is given below.

Lemma 3.1 above provides a closed form expression for the variance of W_i. However, this variance involves the first 4 moments of G, which are usually unknown. Therefore, it is not possible to give a closed form solution for the efficient estimator. Corollary 3.1 provides support that—for large e_i—$\hat{\tau}_2^2$ should be close to the efficient estimator. However, *largeness* is a vague term and it might be valuable to investigate the efficiency of these estimators for real non-random data sets $\{e_i\}$. Now, given any distribution G we are able to compare any linear unbiased estimator to the efficient estimator avoiding any kind of simulation approach. Below, we compare the three estimators $\hat{\tau}_j^2$, for $j = 1, 2, 3$ to the efficient estimator, where the e_i's stem from the two data sets of Example 1 and Example 2, respectively. We choose as heterogeneity distribution G two cases, namely $G_{(a)} = \binom{0.5\ 1.5}{0.5\ 0.5}$ and $G_{(b)} = \binom{0.8\ 0.9\ 1.1\ 1.2}{0.2\ 0.3\ 0.3\ 0.2}$. Here, the notation $G = \binom{\theta_1\ \cdots\ \theta_k}{p_1\ \cdots\ p_k}$ indicates a discrete probability distribution G giving weights $p_1, \ldots, p_k$ to a finite number k of mass points $\theta_1, \ldots, \theta_k$, respectively. Then, the variance of W_i is computed for each $i, i = 1, \ldots, N$ leading to optimal weights $\alpha_i = \operatorname{Var}(W_i)^{-1}$. These optimal weights are compared with the weights used by the three estimators, namely $1/N$, e_i, and e_i^2 by means of scatterplots α_i versus $1/\operatorname{Var}(W_i)$. The closer this relationship is to a straight line with positive slope, the closer is the associated estimator to the efficient one. The results are provided in Fig. 2 and Fig. 3. There is some evidence that $\hat{\tau}_2^2$ is often close to the efficient estimator, since the relationship between the optimal weights and the weights used by this estimator (e_i) appear to be the most linear. This provides some evidence for using $\hat{\tau}_2^2$.

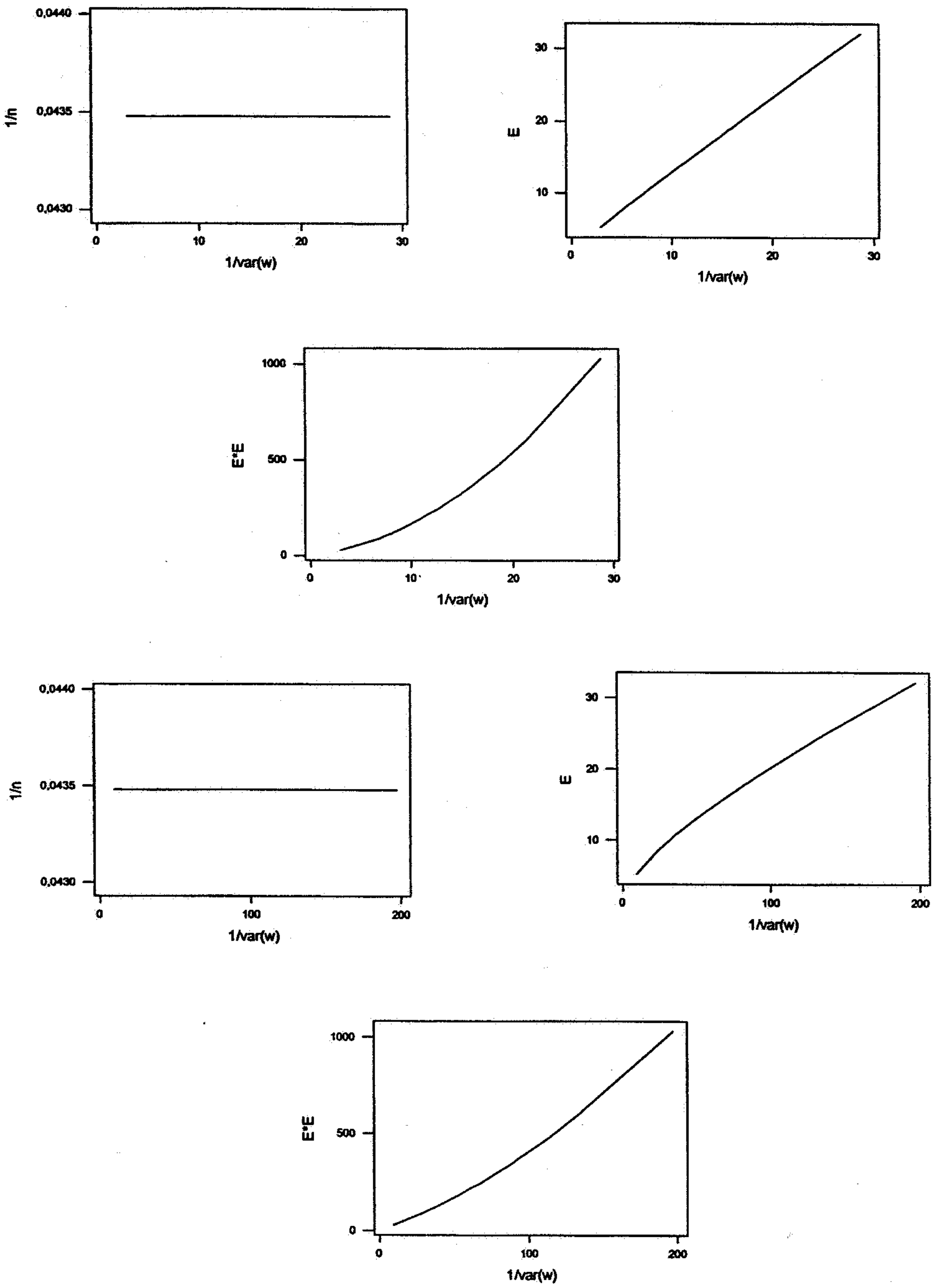

Fig. 2. Scatterplot of the three weighting schemes against $\frac{1}{\mathrm{Var}(W_i)}$ for Hepatits B—data of Example 1 using two different heterogeneity distributions G for computing $\mathrm{Var}(W_i)$; upper page $G_{(a)} = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$, lower page $G_{(b)} = \begin{pmatrix} 0.8 & 0.9 & 1.1 & 1.2 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix}$.

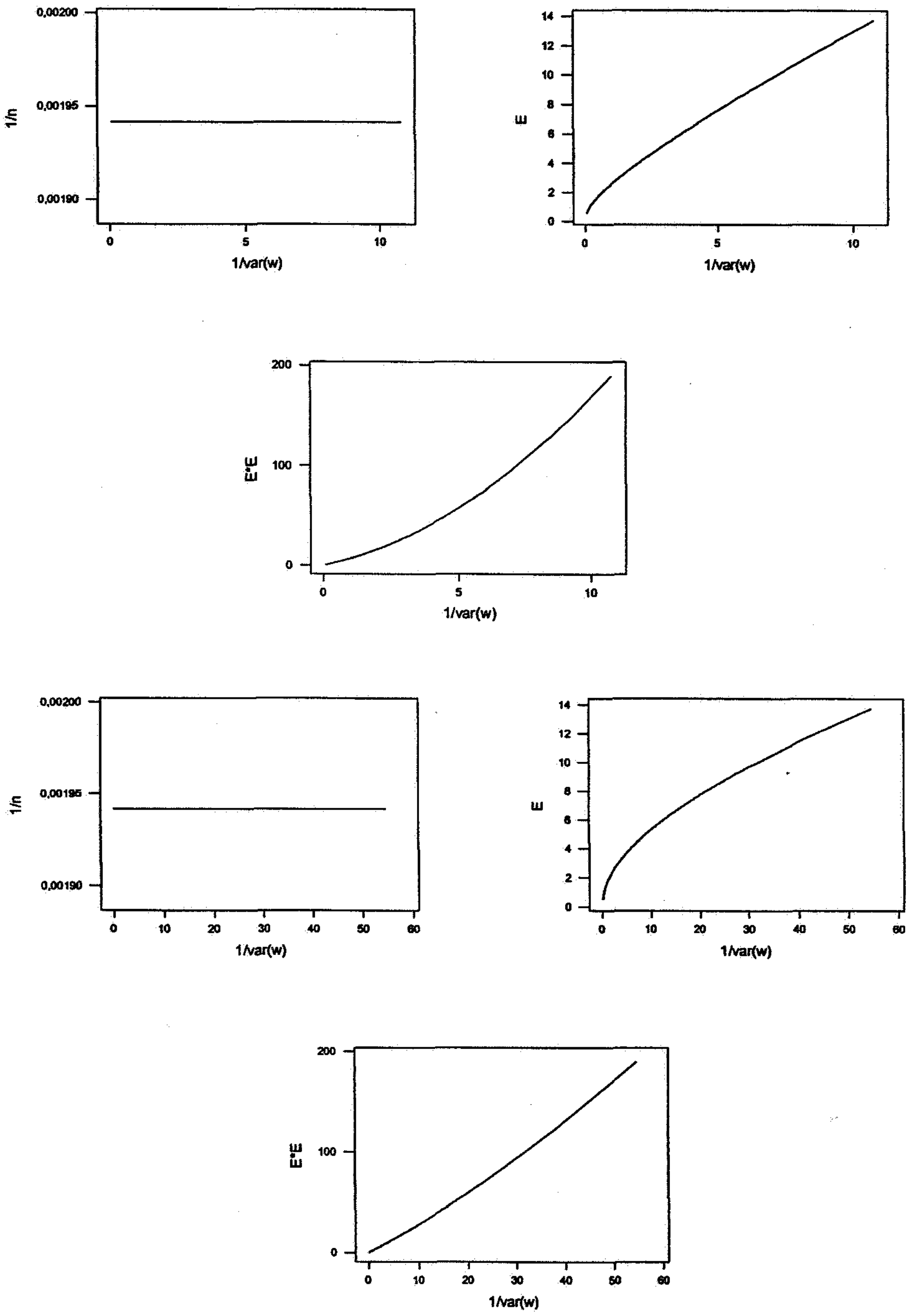

Fig. 3. Scatterplot of the three weighting schemes against $\frac{1}{\mathrm{Var}(W_i)}$ for Perinatal Mortality in the North-West Thames Health Region—data of Example 2 using two different heterogeneity distributions G for computing $\mathrm{Var}(W_i)$; upper page $G_{(a)} = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$, lower page $G_{(b)} = \begin{pmatrix} 0.8 & 0.9 & 1.1 & 1.2 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix}$.

4. Consistency

We are interested in the asymptotic behavior of the estimator $T(W, \alpha)$. For this purpose we require the following two conditions:

(A1) There exists the moment to the power of four for the heterogeneity distribution $G : \mu^{(4)} < \infty$.

(A2) There exist constants $0 < a < A < \infty, 0 < \varepsilon$ such that

$$a \leq \alpha_i \leq A, \qquad \varepsilon \leq e_i \quad \text{for all } i.$$

THEOREM 4.1. *Let* (A1) *and* (A2) *be fulfilled. Then*:

$$T_N(W, \alpha) = \left(\sum_{i=1}^{N} \alpha_i W_i\right) \Big/ \left(\sum_{i=1}^{N} \alpha_i\right) \to \tau^2 \textit{almost surely},$$

in other words, the estimator $T_N(W, \alpha)$ *is* strongly consistent.

PROOF. We have that

$$E(W_i) = \tau^2 \tag{4.1}$$

and under (A1)

$$\begin{aligned} \operatorname{Var}(W_i) &= \mu e_i^{-3} + (2\mu^2 + 7\tau^2)e_i^{-2} + 2(3\mu^{(3)} - 7\mu\tau^2 - 3\mu^3)e_i^{-1} \\ &\quad + 3\mu^4 + \mu^{(4)} - \tau^4 + 6\mu^2\tau^2 - 4\mu\mu^{(3)}. \end{aligned} \tag{4.2}$$

With (A2) it follows, that there exists a finite constant W in such a way, that we have

$$\operatorname{Var}(W_i) \leq W \quad \text{for all} \quad i. \tag{4.3}$$

To obtain W, we have to replace e_i^{-l} by ε^{-l} in (4.2) for $l = 1, 2, 3$. Let us define the following double sequence of random variables:

$$V_i^{(N)} := N \frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j} W_i \quad \text{for} \quad N = 1, 2, \ldots, \; i = 1, \ldots, N. \text{ Note, that}$$

$$\frac{1}{N} \sum_{i=1}^{N} V_i^{(N)} = T_N(W, \alpha). \tag{4.4}$$

For the variables $V_i^{(N)}$ we have that

$$V_1^{(N)}, \ldots, V_N^{(N)} \text{ are independent for all } N, \tag{4.5}$$

$$E(V_i^{(N)}) = N \left(\frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j}\right) \tau^2 \text{ and with this}$$

$$\sum_{i=1}^{N} E(V_i^{(N)}) = N\tau^2, \tag{4.6}$$

$$\operatorname{Var}(V_i^{(N)}) = N^2 \left(\frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j}\right)^2 \operatorname{Var}(W_i).$$

Consequently, there exists a finite constant, say, $\tilde{W} = (A^2/a^2)W$, such that $\text{Var}(V_i^{(N)}) \leq \tilde{W}$ for all $i = 1, \dots, N$ and all $N \geq 1$. With this, it follows that

$$\lim_{N\to\infty} \sum_{i=1}^{N} \frac{\text{Var}(V_i^{(N)})}{i^2} \leq \tilde{W} \lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{i^2} = \tilde{W} \sum_{i=1}^{\infty} \frac{1}{i^2} = \tilde{W}\frac{\pi^2}{6} < \infty. \tag{4.7}$$

According to the strong law of large numbers by *Kolmogorov*, it follows from (4.5) and (4.7)

$$\frac{1}{N}\sum_{i=1}^{N} V_i^{(N)} - \frac{1}{N}\sum_{i=1}^{N} E(V_i^{(N)}) \to 0 \text{ almost surely.}$$

Because of (4.4) and (4.6) this is equivalent to $T_N(W, \alpha) \to \tau^2$ almost surely.

As a consequence we note that $\hat{\tau}_2^2$ and $\hat{\tau}_3^2$ are *strongly* consistent, if there exist positive bounds e, E such that $0 < e \leq e_i \leq E$ for all i. For $\hat{\tau}_1^2$ consistency follows from the fact that in this case we have $V_i^{(N)} = W_i$ as well as (4.3), leading to the inequality (4.7) with W instead of $\tilde{W}$.

5. Estimating heterogeneity mean and variance simultaneously

In many situations, however, it is not appropriate to assume that μ is known. Therefore, we have to replace μ in W_i by some estimate $\hat{\mu}$ leading to

$$W_i(\hat{\mu}) = \frac{(Y_i - e_i\hat{\mu})^2 - e_i\hat{\mu}}{e_i^2}. \tag{5.1}$$

Although only linear unbiased estimators $\hat{\mu}$ might be considered for μ, $W_i(\hat{\mu})$ is *not* necessarily unbiased for τ^2. This fact will cause a bias in $T(W(\hat{\mu}), \alpha)$. The bias will depend on the form of $T(W(\hat{\mu}), \alpha)$ as well as on $\hat{\mu}$ itself. Typically, two mean estimators are considered: the arithmetic mean $\hat{\mu}_1 = \frac{1}{N}\sum_i Y_i/e_i$ and the pooled mean $\hat{\mu}_2 = \frac{\sum_i Y_i}{\sum_i e_i}$. In Böhning (2000), the estimators

$$\hat{\tau}_1^2(\hat{\mu}_j) = \frac{1}{N-1}\left[\sum_{i=1}^{N}(Y_i - e_i\hat{\mu}_j)^2/e_i^2\right] - \hat{\mu}_j\frac{1}{N}\sum_{i=1}^{N}\frac{1}{e_i} \tag{5.2}$$

for $j = 1, 2$ were considered. It was shown that $\hat{\tau}_1^2(\hat{\mu}_1)$ is *unbiased* whereas $\hat{\tau}_1^2(\hat{\mu}_2)$ is biased. This property (unbiasedness) might be one reason to consider $\hat{\tau}_1^2(\hat{\mu}_1)$ at all. For the Hepatitis B data of Berlin we find the results as given in Table 2.

In the light of Section 3, attention is given to the estimator $\hat{\tau}_2^2(\hat{\mu}_j)$ for $j = 1, 2$. It is possible to provide exact expressions for their biases.

Table 2. Estimates of the mean and variance of the SMRs and $\hat{\tau}_1^2$ for Hepatitis B cases in the 23 city regions of Berlin (1995).

Estimator	$\hat{\mu}$	$\widehat{\text{Var}(SMRs)}$	$\hat{\tau}_1^2$	$\hat{\tau}_1^2/\widehat{\text{Var}(SMRs)}$
simple mean	0.9751	0.6214	0.5489	0.883
pooled mean	1.0188	0.6234	0.5470	0.877

THEOREM 5.1. *Let* $\hat{\tau}_2^2(\hat{\mu}_j) = \frac{\sum_{i=1}^N (Y_i - e_i\hat{\mu}_j)^2/e_i - \hat{\mu}_j N}{\sum_{i=1}^N e_i}$ *for* $j = 1, 2$. *Then:*

$$E(\hat{\tau}_2^2(\hat{\mu}_1)) = \left(1 - \frac{1}{n}\right)\tau^2 + \left(\frac{1}{n^2}\sum_i \frac{1}{e_i} - 2\frac{1}{\sum_i e_i}\right)\mu \tag{5.3}$$

$$E(\hat{\tau}_2^2(\hat{\mu}_2)) = \left(1 - \frac{\sum_i e_i^2}{(\sum_i e_i)^2}\right)\tau^2 - \frac{1}{\sum_i e_i}\mu. \tag{5.4}$$

The proof of this theorem is straightforward.

5.1 *Perinatal mortality in the North-West Thames Health Region*

For the data of Example 2, the following values of the biasing constants have been found: $(1 - \frac{1}{n}) = 0.998058$, $(1 - \frac{\sum_i e_i^2}{(\sum_i e_i)^2}) = 0.997376$, and $(\frac{1}{n^2}\sum_i \frac{1}{e_i} - 2\frac{1}{\sum e_i}) = -0.000206377$, $\frac{1}{\sum_i e_i} = 0.000487571$. This example illustrates that the amount of bias involved in expressions (5.3) or (5.4) respectively might be very small.

6. Discussion

The results of this paper can be used for several applications. It was mentioned earlier that the crude *SMR* has several disadvantages including some instability problems for small sample size applications (Lawson *et al.* (1999)). Typical examples are disease mapping and meta-analysis (Böhning (2000)). In these cases, it is more appropriate to use an empirical Bayes estimate of the *SMR*. Often this takes the form $\frac{Y_i + \mu^2/\tau^2}{e_i + \mu/\tau^2}$. It can be shown that this is the *linear Bayes* estimator with respect to the euclidean loss function and it is also the posterior mean if the prior is assumed to be a Gamma distribution (and $Y_i \sim Po(\theta e_i)$) (For details see Böhning (2000)). Clearly, μ and τ^2 need to be replaced by estimates and those that are proposed in this paper might be used for this purpose.

The advantage of the proposed estimators lies in their simple and non-iterative nature. Nevertheless, it should be pointed out that there are many other estimators leading to iterative solutions. One should mention the moment-estimators suggested by Breslow (1984) and Clayton and Kaldor (1987), or the pseudo-maximum-likelihood estimator suggested by Pocock *et al.* (1981), and Breslow (1984). These estimators have been well motivated when they were suggested, and they might be superior in their efficiency to the estimators proposed here. However, a thorough investigation and comparison of these estimators, either in terms of comparing these *iterative* estimators to each other, or comparing the *iterative* estimators to the *non-iterative* estimators suggested here, has not been done yet and is expected to be dealt with in future research.

Acknowledgement

This research is done under support of the German Research Foundation.

REFERENCES

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1990). *Statistical Modelling in GLIM*, Clarendon Press, Oxford.

Böhning, D. (1994). A note on a test for Poisson overdispersion, *Biometrika* **81**, 418–419.

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping, and Others*, Chapman & Hall / CRC, Boca Raton.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models, *Applied Statistics*, **33**, 38–44.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics*, **43**, 671–681.

Collet, D. (1991). *Modelling Binary Data*, Chapman & Hall / CRC, Boca Raton.

Haight, F. A. (1967). *The Handbook of the Poisson Distribution*, Wiley, New York.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F. and Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health*, Wiley, New York.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical Bayes estimators, *Applied Statistics*, **40**, 283–294.

Martuzzi, M. and Elliot, P. (1996). Empirical Bayes estimation of small area prevalence of non-rare conditions, *Statistics in Medicine*, **15**, 1867–1873.

Martuzzi, M. and Hills, M. (1995). Estimating the degree of heterogeneity between event rates using likelihood, *American Journal of Epidemiology*, **141**, 369–374.

Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M. and Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data, *International Statistical Review*, **64**, 89–118.

Pocock, S. J., Cook, D. G. and Beresford, S. A. A. (1981). Regression of area mortality rates on explanatory variables: What weighting is appropriate?, *Applied Statistics*, **30**, 286–295.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models, *Applied Statistics*, **31**, 144–148.