Biostatistics (2004), **5**, 1, *pp.* 61–74
Printed in Great Britain

A comparison of non-iterative and iterative estimators of heterogeneity variance for the standardized mortality ratio

DANKMAR BÖHNING*
Biometry and Epidemiology, Institute for International Health, Joint Center for Health Sciences and Humanitites, Free University Berlin/Humboldt University at Berlin, Haus 562, Fabeckstr. 60-62, 14195 Berlin, Germany
boehning@zedat.fu-berlin.de

JESUS SAROL
Department of Epidemiology and Biostatistics, College of Public Health, University of the Philippines at Manila, Manila, Philippines

SASIVIMOL RATTANASIRI
Clinical Epidemiology Unit, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

CHUKIAT VIWATWONGKASEM
Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand

ANNIBALE BIGGERI
Department of Statistics, University of Florence, 50134 Florence, Italy

SUMMARY

This paper continues work presented in Böhning *et al*. (2002b, *Annals of the Institute of Statistical Mathematics* **54**, 827–839, henceforth BMSRB) where a class of non-iterative estimators of the variance of the heterogeneity distribution for the standardized mortality ratio was discussed. Here, these estimators are further investigated by means of a simulation study. In addition, iterative estimators including the Clayton–Kaldor procedure as well as the pseudo-maximum-likelihood (PML) approach are added in the comparison. Among all candidates, the PML estimator often has the smallest mean square error, followed by the non-iterative estimator where the weights are proportional to the external expected counts. This confirms the theoretical result in BMSRB in which an asymptotic efficiency could be proved for this estimator (in the class of non-iterative estimators considered). Surprisingly, the Clayton–Kaldor iterative estimator (often recommended and used by practitioners) performed poorly with respect to the MSE. Given the widespread use of these estimators in disease mapping, medical surveillance, meta-analysis and other areas of public health, the results of this study might be of considerable interest.

Keywords: Comparative simulation study; DerSimonian–Laird estimator; Heterogeneity variance estimators; Iterative and non-iterative estimators; Moment estimator; Population heterogeneity; Standardized mortality ratio.

*To whom correspondence should be addressed.

1. INTRODUCTION

In a variety of biometric applications, population heterogeneity occurs. In particular, this is the case if there is good reason to model the variable of interest Y through a density of parametric form $p(y|\theta)$ with a scalar parameter θ. For a given subpopulation, the density $p(y|\theta)$ might be very suitable, but a fixed value of θ is not able to cover the whole population of interest. In these situations we speak of extra heterogeneity, which might be caused by unobserved covariates or clustered observations, such as herd clustering in estimating animal infection rates. This situation has been discussed in detail in BMRSB, including a discussion of the background and an illustration of the importance of the problem. An introductory discussion can also 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) as well as Williams (1982); Lee and Nelder (2000); Nelder and Lee (1998) and Lachin (2000, p. 147). The current paper is a follow-up to BMSRB. As before, population heterogeneity means that the parameter of interest, θ, varies in the population, but 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. To be more precise, if θ is itself varying with distribution G and associated density $g(\theta)$, the (unconditional) marginal density of Y is $f(y) = \int_\Theta p(y|\theta)g(\theta)\,\mathrm{d}\theta$. We are interested in the separation of variance into two terms:

$$\mathrm{Var}(Y) = \int_\Theta \mathrm{Var}(Y|\theta)g(\theta)\,\mathrm{d}\theta + \int_\Theta (\mu(\theta) - \mu_Y)^2 g(\theta)\,\mathrm{d}\theta \tag{1}$$

where $\mu(\theta) = E(Y|\theta)$ and $\mu_Y = \int y f(y)\,\mathrm{d}y$ is the marginal mean of Y. Note that $\mu_Y = E_G(\mu(\theta))$. Note that we can also write (1) as

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

In the sequel we will also denote $\mathrm{Var}_G(\mu(\theta))$ by τ_Y^2. Thus, (1) is a partitioning of the variance into components due to the variation in the subpopulation with parameter value θ, averaged over θ, and due to the variance in the heterogeneity distribution G. One can also think of (1) as an analysis-of-variance partition with a latent factor having distribution G. We have to distinguish carefully between *three* distributional schemes when computing moments. For example, $\mathrm{Var}(Y)$ refers to the unconditional or marginal variance and is computed using the marginal density $f(y)$, $\mathrm{Var}(Y|\theta)$ is the *conditional* variance and is computed using the conditional density $p(y|\theta)$, and $\mathrm{Var}_G(\mu(\theta))$ refers to the distribution G of θ. In BMSRB a class of estimators for τ_Y^2 were suggested without implying knowledge of, or estimating, the latent heterogeneity distribution G. The idea behind all the estimators involves re-writing (1) as

$$\mathrm{Var}_G(\mu(\theta)) = \tau_Y^2 = \mathrm{Var}(Y) - E_G(\sigma^2(\theta)). \tag{2}$$

Replacing $\mathrm{Var}(Y)$ and $E_G\sigma^2(\theta)$ on the right-hand side of (2) by their respective sample estimates we can obtain estimates for τ_Y^2. In the succeeding text, we will use μ as the mean of θ and τ^2 for its variance.

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

In the next section we consider a generalization of this idea to the standardized mortality ratio. In Section 3 we discuss a more general class of linear unbiased estimators of the heterogeneity variance which have been suggested in BMSRB. Section 4 discusses an extension to the case when the mean μ also needs to be estimated, including an appropriately adjusted version of the DerSimonian and Laird (1986)

estimator. Section 5 introduces two iterative estimators. The first one was suggested by Breslow (1984) and later utilized by Clayton and Kaldor (1987) whereas the second one was suggested by Pocock *et al.* (1981) and later discussed by Dean and Lawless (1989). Close connections of these iterative estimators to the non-iterative estimators used in Section 3 are demonstrated. Section 6 discusses the design and analysis of a simulation study to compare these six estimators of the heterogeneity variance.

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 representing a sequence of mortality or morbidity cases. Associated with each replication Y_i is a deterministic quantity e_i which represents an expected count and is usually calculated on the basis of an external reference population.

2.1 *Indirect standardization*

To be more specific about this *indirect method of standardization* (see also Woodward, 1999), let λ_j be the mortality or morbidity rate (cases divided by the number at risk) in age group j of the reference population. Furthermore, let n_{ij} be the number at risk in age group j in replication i in the sample. To think of an example, the replication i could represent a certain region or year of interest for a given country. Then, e_i is simply defined as the expected number of cases in the study population replication i, when the mortality rates of the reference population are valid, namely $e_i = \sum_{j=1}^{J} n_{ij}\lambda_j$, where J is the number of age groups. This form of indirect standardization is called *external*. Sometimes the reference population is not readily available and needs to be constructed. This can be accomplished as follows. Let Y_{ij} denote the number of cases in age group j and replication i of the study population. Then, define $\hat{\lambda}_j = \sum_{i=1}^{N} Y_{ij} / \sum_{i=1}^{N} n_{ij}$ as the ratio of the averages of observed cases and numbers at risk over all replications in age group j. To compute the e_i we proceed as above, namely $e_i = \sum_{j=1}^{J} n_{ij}\hat{\lambda}_j$, where now estimates of the rates of the reference population are used. This form of indirect standardization is called *internal*. The major difference between internal and external indirect standardiziation lies in the fact that in the internal method the mortality rates of the 'reference' population are constructed from the replications in the study population. Although both methods of indirect standardization are quite similar, there is one pecularity of the internal method which needs to be pointed out. It follows directly from the construction that the sum of the observed counts is equal to the sum of the constructed, expected counts, namely

$$\sum_i e_i = \sum_i \sum_j n_{ij} \left[\sum_i Y_{ij} / \sum_i n_{ij} \right] = \sum_j \sum_i Y_{ij} = \sum_i Y_i$$

where $Y_i = \sum_j Y_{ij}$ and, from here, $\sum_i Y_i / \sum_i e_i = 1$. This implies that the marginal mean of the ratios Y_i/e_i is fixed to 1.

2.2 *Definition and properties*

With the help of these numbers e_i one can define the standardized mortality ratio as $\mathrm{SMR}_i = Y_i/e_i$ and its expected value $E(\mathrm{SMR}_i \,|\theta_i) = \theta_i$, for $i = 1, \ldots, N$. Frequently, this sample is associated with N geographic regions or areas, as often arises in *disease mapping*. For an introduction to this field see Böhning (2000) or Lawson *et al.* (1999).

Furthermore, conditional on the value of θ a Poisson distribution is assumed for $Y|\theta$: $p(y_i|\theta, e_i) = \exp(-\theta e_i)(\theta e_i)^{y_i}/y_i!$. For this case, the partition of variance (1) takes the form

$$\mathrm{Var}(Y_i) = E_G(\sigma_i^2(\theta)) + \mathrm{Var}_G(\mu_i(\theta)) = e_i E_G(\theta) + e_i^2 \mathrm{Var}_G(\theta) = e_i\mu + e_i^2\tau^2. \tag{3}$$

2.3 *Simple estimators of the heterogeneity variance*

We write (3) as $E(Y_i - e_i\mu)^2 = e_i\mu + e_i^2\tau^2$, which draws our attention to the variate $W_i = \frac{(Y_i - e_i\mu)^2 - e_i\mu}{e_i^2}$. Since $\mathrm{Var}(Y_i) = E(Y_i - e_i\mu)^2$, it follows from (3) that

$$E(W_i) = \tau^2 \tag{4}$$

One estimate of τ^2 replaces $\mathrm{Var}(Y_i)$ by its 'estimate' $(Y_i - e_i\mu)^2$, solves for τ^2 and then averages over i to give

$$\hat{\tau}_1^2 = \frac{1}{N}\left[\sum_{i=1}^{N}(Y_i - e_i\mu)^2/e_i^2 - \mu\sum_{i=1}^{N}\frac{1}{e_i}\right]. \tag{5}$$

For a second estimate, we first divide by e_i in (3) , then average over i and solve for τ^2 to give

$$\hat{\tau}_2^2 = \frac{\sum_{i=1}^{N}(Y_i - e_i\mu)^2/e_i - \mu N}{\sum_{i=1}^{N} e_i}. \tag{6}$$

A third possibility is to average over i in (3), and then solve for τ^2 to give

$$\hat{\tau}_3^2 = \frac{\sum_{i=1}^{N}(Y_i - e_i\mu)^2 - \mu\sum_{i=1}^{N} e_i}{\sum_{i=1}^{N} e_i^2}. \tag{7}$$

All three estimators coincide if the e_i coincide, and all three are unbiased.

2.4 *The connection to empirical Bayes estimators*

A further motivation for studying estimation of the heterogeneity variance τ^2 is in the context of Bayes estimators for the standardized mortality ratio (SMR). Assuming a Gamma distribution for θ_i, the posterior distribution of θ_i is again a Gamma distribution with posterior mean $\theta_i^{EB} = \frac{Y_i + \mu^2/\tau^2}{e_i + \mu/\tau^2}$, which is called an Empirical Bayes estimator for θ_i. This is also *the best linear Bayes estimator* (for details see Böhning, 2000, pp. 152–157). Empirical Bayes estimators are considered to be superior to the crude SMR, as they help to avoid the occurence of artefacts as described by Clayton and Kaldor (1987), in particular when the number of cases per replication is small. Empirical Bayes estimators allow a direct interpration in terms of population *heterogeneity* in that they coincide with the crude SMR if there is strong heterogeneity (τ^2 large) and coincide with the overall mean if there is no heterogeneity ($\tau^2 = 0$). Thus, one can think of τ^2 as a *smoothing parameter*. To use the Empirical Bayes estimators in practice, one needs to replace the theoretical parameters by estimates, which can be obtained by the methods suggested here.

3. NON-ITERATIVE ESTIMATORS

3.1 *A general class*

The estimators given in the previous section are special cases of a more general class of *linear unbiased estimators* of τ^2:

$$T(W,\alpha)=\frac{\sum_{i=1}^{N}\alpha_i W_i}{\sum_{i=1}^{N}\alpha_i} \tag{8}$$

for any non-random, non-negative numbers $\alpha_1,\alpha_2,\ldots,\alpha_N$. It is easy to verify that $\alpha_i=1/N$, $\alpha_i=e_i$, and $\alpha_i=e_i^2$ lead to the estimators $\hat{\tau}_1^2$, $\hat{\tau}_2^2$ and $\hat{\tau}_3^2$, respectively. The choice $\alpha_i=1/N$ is mentioned in Böhning (2000), whereas $\alpha_i=e_i$ is suggested by Marshall (1991) and mentioned in Lachin (2000, p. 325), and $\alpha_i=e_i^2$ is mentioned in Bautista (1997). Asymptotic properties of these estimators are considered in Moore (1986) and Gourieroux *et al.* (1984). Asymptotic efficiency of $\hat{\tau}_2^2$ has been established in BMSRB.

The estimator $T(W,\alpha)$ requires knowledge of the overall mean μ. This assumption is often satisfied since the SMR_i are indirectly standardized in such a way that $\sum_i Y_i/\sum_i e_i=1$. If the overall mean μ is unknown, it could be estimated by $\sum_i Y_i/\sum_i e_i$.

3.2 *Example 1: hepatitis B in Berlin*

To illustrate the estimators, we consider two examples. Table 1 gives the observed and expected numbers of hepatitis B cases in the 23 city regions of Berlin for the year 1995. Here, we find that $\sum_i Y_i/\sum_i e_i=1.019$ (external method of indirect standardization). A conventional χ^2-test for homogeneity is given by $\chi^2=\sum_i(Y_i-\mu e_i)^2/(\mu e_i)$. If we replace μ by $\hat{\mu}=\sum_i Y_i/\sum_i e_i=1.019$, we get $\chi^2=193.52$, clearly indicating heterogeneity. Assuming for this illustration that μ is fixed we obtain $\hat{\tau}_1^2=0.5205$, $\hat{\tau}_2^2=0.4810$ and $\hat{\tau}_3^2=0.4226$. This indicates rather high heterogeneity since $\widehat{\mathrm{Var}(\mathrm{SMR})}=\frac{1}{N-1}\sum_i(\mathrm{SMR}_i-\overline{\mathrm{SMR}})^2=0.6234$. Note that using a correct estimate of variance leads to an increased length for a 95% confidence interval for μ constructed as $\hat{\mu}\pm 1.96\sqrt{\mathrm{Var}(\hat{\mu})}$ where $\hat{\mu}$ corresponds to the pooled estimator. To see this we look at $\mathrm{Var}(\hat{\mu})=\mu/(\sum_i e_i)+\tau^2\sum_i e_i^2/(\sum_i e_i)^2$. Obviously, the conventional textbook variance formula occurs when τ^2 is set to 0 and μ is estimated by the pooled mean $\sum_i Y_i/\sum_i e_i$, namely $\widehat{\mathrm{Var}(\hat{\mu})}=\sum_i Y_i/(\sum_i e_i)^2$. For details, see Woodward (1999, p. 162). Now, if heterogeneity is present the variance will be underestimated and the associated confidence intervals too small. The length of the correct interval will depend on the way τ^2 is estimated, as illustrated in Table 2. Assessment of the relative merits of different estimators of τ^2 is therefore important.

3.3 *Example 2: perinatal mortality in the North-west Thames health region*

We consider as a second example 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 1986–90. The region consists of 515 small areas. The data (provided by Marco Martuzzzi) are listed in Table 1 of the *supplementary material*. In this case, $\sum_i Y_i=\sum_i e_i=2051$ (internal method of indirect standardization). We find that $\hat{\tau}_1^2=-0.0273$, which we replace by zero, $\hat{\tau}_2^2=0.0168$ and $\hat{\tau}_3^2=0.0370$. There is small heterogeneity present in the data which is indicated by the ratio $\hat{\tau}_j^2/\widehat{\mathrm{Var}(\mathrm{SMR})}$, where $\widehat{\mathrm{Var}(\mathrm{SMR})}=\frac{1}{N-1}\sum_i(\mathrm{SMR}_i-\overline{\mathrm{SMR}})^2=0.6058$.

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	14	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

Table 2. *Estimators of heterogeneity variance, variance of pooled mean with 95% confidence interval for μ*

j	$\hat{\tau}_j^2$	$\widehat{\mathrm{Var}(\hat{\mu})}$	95% Confidence interval
1	0.5205	0.0286	(0.6875, 1.3505)
2	0.4810	0.0267	(0.6990, 1.3389)
3	0.4226	0.0238	(0.7169, 1.3211)
Conventional	0	0.0028	(0.9149, 1.1231)

3.4 *An estimator of heterogeneity variance according to DerSimonian–Laird*

In this section the DerSimonian–Laird estimator is considered in its general form. Suppose that a random sample $x_1, x_2, \ldots, x_k$ of size k is available with associated variances $v_1^2, v_2^2, \ldots, v_k^2$. In a meta-analytic setting this sample would represent a collection of k independent studies for which within study j a statistic x_j is measured with standard error v_j. Then the following result holds:

$$E(\chi^2) = (k-1) + \tau^2 \left(\sum_{i=1}^{k} w_i - \sum_{i=1}^{k} w_i^2 / \sum_{i=1}^{k} w_i \right) \tag{9}$$

where $w_i = 1/v_i^2$, $\chi^2 = \sum_{i=1}^{k} w_i (x_i - \hat{\mu})^2$, and $\hat{\mu} = \sum_{i=1}^{k} w_i x_i / \sum_{i=1}^{k} w_i$. The proof is along the lines of the proof given in Böhning (2000) where the simpler case $v^2 = \sigma^2$ is considered. Equating the expected value (9) to the empirical observed χ^2-value leads to the *moment estimator*

$$\hat{\tau}^2_{\mathrm{DSL}} = \frac{\chi^2 - (k-1)}{\sum_{i=1}^{k} w_i - \sum_{i=1}^{k} w_i^2 / \sum_{i=1}^{k} w_i}. \tag{10}$$

Note that the estimator (10) is *unbiased* by construction. This result is unaffected by the distributional properties of the χ^2-statistic. It was originally developed by DerSimonian and Laird (1986) for the case of Normally distributed effect measures with *known* variances. Because of this tradition we refer to the estimator (10) as the DerSimonian–Laird (DSL) estimator.

3.4.1 *The case of SMRs.* Suppose that $x_i = Y_i/e_i$ where e_i is known, Y_i follows a Poisson distribution with parameter $e_i\theta$ and θ follows a distribution G with mean μ and variance τ^2. Then the conditional, population averaged variance $\int_\theta \mathrm{Var}(\mathrm{SMR}_i\,|\theta)\,\mathrm{d}\theta = E_G\,\mathrm{Var}(\mathrm{SMR}_i\,|\theta)$ of $x_i = \mathrm{SMR}_i$ is

$$v_i^2 = E_G\,\mathrm{Var}(\mathrm{SMR}_i\,|\theta) = \mu/e_i. \tag{11}$$

When estimating μ in (11) one could use either the pooled or simple mean estimate. Details of using the DSL estimator in the context of SMR data can be found in Böhning *et al.* (2002a).

4. Estimating heterogeneity mean and variance

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

$$W_i(\hat{\mu}) = \frac{(Y_i - e_i\hat{\mu})^2 - e_i\hat{\mu}}{e_i^2}. \tag{12}$$

Although we might consider only linear unbiased estimators $\hat{\mu}$ for μ, $W_i(\hat{\mu})$ is *not* necessarily unbiased for τ^2. This will lead to 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 $\hat{\mu}_1 = N^{-1}\sum_i Y_i/e_i$ and the pooled mean $\hat{\mu}_2 = \sum_i Y_i/\sum_i e_i$. In Böhning (2000) the estimators

$$\hat{\tau}_1^2(\hat{\mu}_j) = \frac{1}{N-1}\left[\sum_{i=1}^{N}(Y_i - e_i\hat{\mu}_j)^2/e_i^2\right] - \hat{\mu}_j\frac{1}{N}\sum_{i=1}^{N}\frac{1}{e_i} \tag{13}$$

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, hence $\hat{\tau}_1^2(\hat{\mu}_1)$ may well be preferred.

Let us consider the efficient estimation of μ. Consider the sample of SMR values $Y_1/e_1, Y_2, e_2, \ldots, Y_N/e_N$. According to Section 2 we have that $\mathrm{Var}(\mathrm{SMR}_i) = \mu/e_i + \tau^2$. Therefore, the best linear unbiased estimator for μ is given as

$$\hat{\mu} = \frac{\sum_{i=1}^{N}(\mu/e_i + \tau^2)^{-1}\,\mathrm{SMR}_i}{\sum_{i=1}^{N}(\mu/e_i + \tau^2)^{-1}} \tag{14}$$

which can also be derived from the quasi-likelihood approach. Note that if $\tau^2 = 0$, (14) coincides with $\hat{\mu}_2 = \sum_i Y_i/\sum_i e_i$, whereas as $\tau^2 \to \infty$, (14) approaches $\hat{\mu}_1 = \frac{1}{N}\sum_i \mathrm{SMR}_i$. Otherwise, (14) will depend on the value of τ^2. In the following sections we consider other ways to estimate τ^2.

5. Iterative estimators

5.1 *Moment estimators for* τ^2

Breslow (1984) suggested a moment estimator for τ^2 based on the chi-square statistic. This was later used by Clayton and Kaldor (1987). Though they started out with a maximum likelihood approach based upon a Poisson–Gamma model, they used the moment estimator in the estimation algorithm, probably because of the numerical complexities involved in the Gamma-function. In the light of DerSimonian and Laird (1986), define

$$\chi^2 = \sum_{i=1}^{N}\frac{(Y_i - \mu e_i)^2}{\mu e_i + \tau^2 e_i^2}$$

and equate it to N. This leads to an implicit equation for τ^2 as given by Clayton and Kaldor (1987); namely

$$\tau^2 = \frac{1}{N}\sum_i \left(\frac{Y_i + \mu^2/\tau^2}{e_i + \mu/\tau^2} - \mu\right)^2 \left(1 + \frac{\mu}{\tau^2 e_i}\right). \tag{15}$$

Equation (15) can be given in equivalent form as a linear combination of the W_i, namely

$$\tau^2 = \frac{\sum_i \alpha_i(\tau^2) W_i}{\sum_i \alpha_i(\tau^2)} \tag{16}$$

where $\alpha_i(\tau^2)^{-1} = \mu/e_i + \tau^2$. Note that equations (14)–(16) are fixed-point equations and can be used constructively to find $\hat{\mu}$ and $\hat{\tau}^2$. In particular, let $\hat{\tau}^2_{MO,n} = \sum_i \alpha_i(\hat{\tau}^2_{MO,n-1}) W_i / \sum_i \alpha_i(\hat{\tau}^2_{MO,n-1})$ for $n = 1, 2, \ldots$, and any $\hat{\tau}^2_{MO,0} \geqslant 0$, be a sequence generated by (16). Any estimator fulfilling (15) will be denoted by $\hat{\tau}^2_{MO}$, and under regularity conditions as discussed in Section 5.4, a sequence $\hat{\tau}^2_{MO,n}$ will converge to a solution of (16).

5.2 *Pseudo-maximum likelihood*

Another method was used by Pocock *et al.* (1981), Dean and Lawless (1989) and is mentioned also in Breslow (1984). The idea is to treat Y_i as if it were Normally distributed with mean μe_i and variance $\mu e_i + \tau^2 e_i^2$. The associated log-likelihood is proportional to

$$L(\tau^2) = -\sum_i \log(\mu e_i + \tau^2 e_i^2) - \sum_i \frac{(Y_i - \mu e_i)^2}{\mu e_i + \tau^2 e_i^2}$$

and the associated score-equation leads to

$$\frac{\partial L}{\partial \tau^2} = -\sum_i \frac{e_i^2}{\mu e_i + \tau^2 e_i^2} + \sum_i \frac{(Y_i - \mu e_i)^2 e_i^2}{(\mu e_i + \tau^2 e_i^2)^2} = 0$$

which again can be written as a weighted sum of the W_i:

$$\tau^2 = \frac{\sum_i \alpha_i(\tau^2) W_i}{\sum_i \alpha_i(\tau^2)} \tag{17}$$

where $\alpha_i(\tau^2)^{-1} = (\mu/e_i + \tau^2)^2$. Equation (17) is again a fixed-point equation which can be used constructively, in the same way as has been described in the previous section. Any estimator fulfilling (17) will be denoted by $\hat{\tau}^2_{PML}$. There are similarites between this procedure and the marginal likelihood method of Hardy and Thompson (1996), in particular the iterative procedures given by their equations (8) and (9). However, whereas in Hardy and Thompson the variance of each replication is assumed to be *known*, in our case the variance in each study has a specific structure, namely μ/e_i, which avoids such a crucial assumption. In our context, e_i is known whilst μ is independent of the replication, and completely specified if internal indirect standardization is used. Otherwise, μ can simply be estimated using the pooled or simple mean.

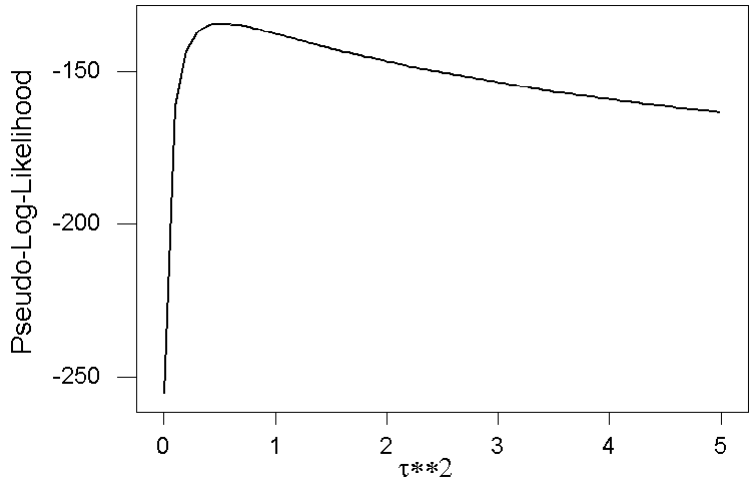

Fig. 1. Pseudo-log-likelihood for hepatitis data of Berlin.

5.3 *Non-iterative as one-step-iterative estimators*

The connection between the iterative estimators defined by (15)–(17) and the class of non-iterative estimators $\tau^2 = \sum_i \alpha_i W_i / \sum_i \alpha_i$ is surprisingly close. Consider the moment estimator of Breslow–Clayton–Kaldor, implicitly defined by $\tau^2 = \sum_i \alpha_i(\tau^2) W_i / \sum_i \alpha_i(\tau^2)$, where $\alpha_i(\tau^2) = (\mu/e_i + \tau^2)^{-1}$. If we choose 0 as initial value for τ^2, we have that $\hat{\tau}^2_{MO,1} = \sum_i e_i W_i / \sum_i e_i$ which coincides with $\hat{\tau}^2_2$. Similarly, the first step of the PML procedure (17) leads to $\hat{\tau}^2_{PML,1} = \sum_i e_i^2 W_i / \sum_i e_i^2$, which is exactly $\hat{\tau}^2_3$. Finally, one might wonder in which situation $\hat{\tau}^2_1$ will occur as a special case of $\hat{\tau}^2_{MO}$ or $\hat{\tau}^2_{PML}$. In fact, this occurs as a limiting case when τ^2 is large relative to μ/e_i.

5.4 *Convergence problems for the iterative estimators*

When using procedures which iteratively construct the estimator of interest several issues should be considered. The first issue is that of an appropriate *starting value*. In our case, all procedures were started with $\tau^2 = 0$. One justification can be seen in the previous section where it was shown that with this starting value the first step coincides with good estimators from the non-iterative family. The second issue is more crucial: when should we stop the iteration? Conventionally, the iteration is stopped when two consecutively generated estimates are close to each other, where closeness is defined by some value ϵ, in our case $\epsilon = 0.000\,01$. Typically, both iterative procedures converge very quickly, in most cases reaching the stopping criterion in less than five steps. An example is given in Figure 1, which shows the pseudo-log-likelihood for the hepatitis B data of Berlin. However, occasionally the iteration slows down, especially in cases where the pseudo-log-likelihood is rather flat, as in the top panel of Figure 2. Here, the iteration process might need several hundred iterations. A more disturbing and frequently occurring problem is *non-convergence*, for which an example is provided in the bottom panel of Figure 2. The reason is here that in the population used in the simulation the heterogeneity variance τ^2 is small, the pseudo-log-likelihood has no point of stationarity and the maximum occurs at the boundary point $\tau^2 = 0$; typically the iteration jumps back and forth between two points. This case can be diagnosed easily by investigating the first derivative of the pseudo-log-likelihood at 0; if it is negative the PML estimator

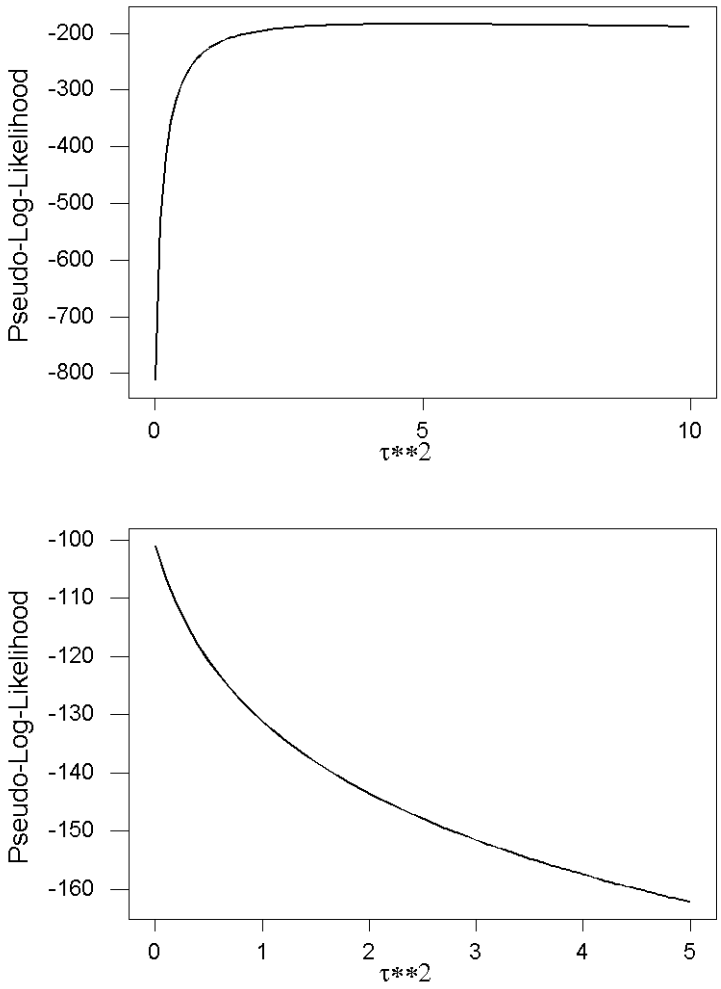

Fig. 2. Pseudo-log-likelihood for simulated data, expected counts from hepatitis data of Berlin; data set 1: unique mode, but flat likelihood, slow convergence of iteration procedure (top); Data Set 2: maximum occurs at boundary and and log-likelihood has no stationarity point (bottom).

is taken to be 0. Similiar considerations can be undertaken for the fixed-point iteration of the Clayton–Kaldor procedure. These simple diagnostic devices help to avoid pitfalls during the iteration process. Though these diagnostic tools do not ensure convergence of the iterative procedure for every potential constellation of the data, they provide some protection against *typical* causes of non-convergence. In fact, in the simulation study described in Section 6 the iterative procedure has been executed several million times with no cases of non-convergence.

5.5 *An illustration of all estimators*

Here all six estimators, namely the four *non-iterative* estimators $\hat{\tau}_1^2$, $\hat{\tau}_2^2$, $\hat{\tau}_3^2$, $\hat{\tau}_{\text{DSL}}^2$ and the two *iterative* estimators $\hat{\tau}_{\text{MO}}^2$ and $\hat{\tau}_{\text{PML}}^2$, are illustrated for the hepatitis B data of Berlin.

Table 3. *Six estimators of heterogeneity variance illustrated for the hepatitis B data of Berlin*

	Non-iterative				Iterative	
μ estimated by	$\hat{\tau}_1^2$	$\hat{\tau}_2^2$	$\hat{\tau}_3^2$	$\hat{\tau}_{DSL}^2$	$\hat{\tau}_{MO}^2$	$\hat{\tau}_{PML}^2$
simple	0.5219	0.4857	0.4301	0.5118	0.5207	0.5189
pooled	0.5205	0.4810	0.4226	0.5090	0.5187	0.5163

There is considerable variation in the six estimators. The two iterative estimators appear to be close to the DerSimonian–Laird estimator and $\hat{\tau}_1^2$, whereas the two non-iterative estimators $\hat{\tau}_2^2$ and $\hat{\tau}_3^2$ appear to be lower in value. Note that estimating μ with the pooled mean (no heterogeneity) or the simple mean (large heterogeneity) seems to have a minor effect on all of the variance estimators.

6. Simulation study

Though empirical data are useful in illustrating the behaviour of estimators, they are not helpful in evaluating their statistical properties. For this purpose a simulation study has been undertaken. The objective of the simulation study is to compare the four *non-iterative* estimators $\hat{\tau}_1^2$, $\hat{\tau}_2^2$, $\hat{\tau}_3^2$, $\hat{\tau}_{DSL}^2$ and the two *iterative* estimators $\hat{\tau}_{MO}^2$ and $\hat{\tau}_{PML}^2$ with respect to bias, variance and mean square error (MSE). All six estimators will depend on the value of μ, which can be taken to be *known* (internal indirect standardization) or unknown. For the latter, we consider the *pooled* mean estimator $\hat{\mu}_2 = \sum_i Y_i / \sum_i e_i$ and the *simple* mean estimator $\hat{\mu}_1 = N^{-1} \sum_i Y_i / e_i$.

6.1 *Design*

We have considered a mixture of two Poisson distributions. The mixing distributions gives weight p to θ_1 and weight $1 - p$ to θ_2. Consequently, the marginal density is given by

$$f(y|e) = pPo(y|\theta_1, e) + (1 - p)Po(y|\theta_2, e), y = 0, 1, 2, \ldots. \tag{18}$$

where e is the number of expected cases associated with y. We fixed the parameter for the first component, $\theta_1 = 1$, leading to the mean and variance of the mixing distribution as

$$\mu = p\theta_1 + (1 - p)\theta_2 = p + (1 - p)\theta_2$$

$$\tau^2 = p(\theta_1 - \mu)^2 + (1 - p)(\theta_2 - \mu)^2 = (1 - p)p(1 - \theta_2)^2, \tag{19}$$

but allowed the mean μ of the mixing distribution to take values 1.5, 2.0, 3.0, 4.0 and 5.0, and the variance τ^2 to take values 0.1, 0.2, 0.5, 1.0 and 2.0. For each pair of values of μ and τ^2, and with $\theta_1 = 1$, the corresponding values of p and θ_2 are

$$p = \frac{\tau^2}{(\mu - 1)^2 + \tau^2} \tag{20}$$

$$\theta_2 = (\mu - p)/(1 - p). \tag{21}$$

The SMR is computed by dividing the observed number of cases by the expected number, i.e. $\text{SMR}_i = y_i/e_i$. The e_i are treated as fixed quantities and are usually computed based on an external reference

population. For this study, the e_i came from three different sources: (1) the expected cases for the hepatitis B data in 23 city regions of Berlin in 1995 (Table 1), (2) expected numbers set uniformly in steps from 1.05 to 11, giving exactly 200 e_i; and (3) the expected numbers of a data set of perinatal mortality in the North-west Thames health region, England in the period 1986–90 on the basis of 515 small areas as discussed by Martuzzi and Hills (1995). For each set of e_i, we generated a corresponding set of observed frequencies y_i using specific parameter combinations of μ and τ^2 as outlined in the next section.

Given the parameters θ_1, θ_2 and p, we then proceeded to obtain a simulation of the observed frequencies. For each e_i, we first generated a random number, u, from the uniform distribution $U(0, 1)$. If $u < p$, then we generated a random number y_i from a Poisson distribution with parameter $\theta_1 e_i$; otherwise, we generated y_i from a Poisson distribution with parameter $\theta_2 e_i$. After the sample is completed μ and τ^2 are estimated using the different methods as discussed in the next section.

There are three situations for the mean, namely the simple mean $\hat{\mu} = N^{-1} \sum_{i=1}^{N} y_i/e_i$, the pooled mean $\hat{\mu} = \sum_{i=1}^{N} y_i / \sum_{i=1}^{N} e_i$, and the case of known μ. In combination with the six estimators for τ^2 conditional on μ, this leads to 18 estimators of τ^2. This process is replicated 10 000 times, and the MSE and bias are calculated as

$$\text{Bias} = \frac{1}{10\,000} \sum_{i=1}^{10\,000} \hat{\tau}_i^2 - \tau^2 \tag{22}$$

$$\text{Variance} = \frac{1}{10\,000} \sum_{i=1}^{10\,000} (\hat{\tau}_i^2 - \bar{\tau}^2)^2 \tag{23}$$

$$\text{MSE} = \frac{1}{10\,000} \sum_{i=1}^{10\,000} (\hat{\tau}_i^2 - \tau^2)^2. \tag{24}$$

6.2 *Results*

An overall picture is provided in Figure 3 where average rankings are taken over all replications and over all populations considered. For each of the 25 heterogeneity populations studied the six estimators have been ranked according to their MSE. Since there are many populations some overall measure needs to be considered: here, the mean rank was chosen. The best estimator is the moment method of Clayton–Kaldor for the hepatitis B e_i, the non-iterative estimator $\hat{\tau}_2^2$ for the perinatal mortality e_i and for the artificial e_i. The worst estimator is $\hat{\tau}_3^2$ for the hepatitis B e_i, $\hat{\tau}_1^2$ for the perinatal mortality e_i and for the artificial e_i. Note that the Clayton–Kaldor estimator (ranked 1 for the hepatitis B set) is ranking only on 4 for the two other constellations of the e_i. It appears that the PML estimator is doing quite well independent of the constellations of the e_i: it is ranked 2 in all three constellations. This impression is confirmed to a larger extent when only populations with large heterogeneity are considered. Here the PML methods ranks 1 in two constellations, and ranks 2 in the other constellation. More details can be found in the *supplementary material*.

When only the four non-iterative estimators are compared, we find that the Marshall estimator $\hat{\tau}_2^2$ is perfoming well: it is ranked 1 for two constellations, only for the hepatitis B set of e_i is it outperformed by $\hat{\tau}_1^2$. This is not different for populations with large heterogeneity. Note also that the DerSimonian–Laird estimator $\hat{\tau}_{DSL}^2$ ranks directly behind the Marshall estimator $\hat{\tau}_2^2$ in all three constellations of e_i. Next, we consider the dependence of the ranking according to the estimation method for μ: it is either assumed that μ is known, or estimated by the simple mean, or estimated by the pooled mean. The results do not

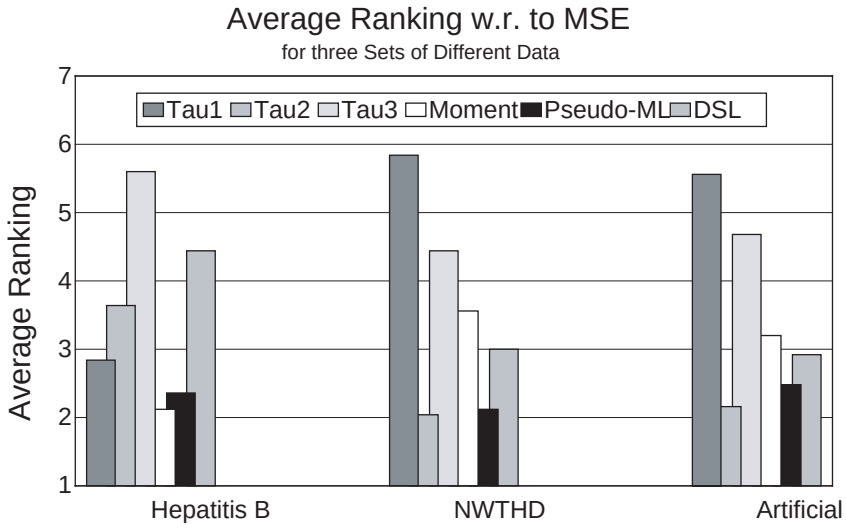

Fig. 3. Average ranking of the six estimators (averaged over all replications and all populations): $\hat{\tau}_1^2$, $\hat{\tau}_2^2$, $\hat{\tau}_3^2$, $\hat{\tau}_{MO}^2$, $\hat{\tau}_{PML}^2$, $\hat{\tau}_{DSL}^2$ (from the left to the right).

change qualitatively. This seems to justify a summary averaged over the methods of estimation for μ, as presented in Figure 3. More details on this point can be found in the *supplementary material*.

If the MSE is classified according to the value of τ^2, we can see that the relative performance of $\hat{\tau}_1^2$ improves with increasing heterogeneity. The reason is that in the case of large heterogeneity the weights $(\mu/e_i + \tau^2)^{-1}$ which combine the W_i become more similar, thus making $\hat{\tau}_1^2$ close to the iterative PML procedure. Again, more details on this point can be found in the *supplementary material*.

Which procedure should be chosen? Amongst the iterative procedures, there appears to be evidence to recommend the PML approach. This iteration should be accompanied by diagnostics for a maximum on the boundary (negative derivative at 0), which could be done on initialization when the starting value of 0 is used. Alternatively, one might consider a non-iterative estimator, even for reasons of choosing a good initial value. Here, the Marshall estimator appears to have the best performance. The simulation study also provides evidence that there is not too much loss in efficiency if this non-iterative estimator is used in comparison to the PML estimator.

ACKNOWLEDGEMENTS

This research was done under support of the *German Research Foundation.* The authors would like to express their gratitude to an unknown referee as well as to the editor Peter Diggle.

REFERENCES

AITKIN, M., ANDERSON, D., FRANCIS, B. AND HINDE, J. (1990). *Statistical Modelling in GLIM*. Oxford: Clarendon.

BAUTISTA, O. M. (1997). Analysis of overdispersed Poisson count, *Doctoral Dissertation*, UMI, Ann Arbor.

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*. Boca Raton, FL: Chapman and Hall/CRC.

BÖHNING, D., DIETZ, E., SCHLATTMANN, P., VIWATWONKASEM, C. AND BIGGERI, A. (2002a). Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics* **3**, 445–457.

BÖHNING, D., MALZAHN, U., SAROL, J., RATTANASIRI, S. AND BIGGERI, A. (2002b). Efficient non-iterative and nonparametric estimation of heterogeneity variance for the standardized mortality ratio. *Annals of the Institute of Statistical Mathematics* **54**, 827–839.

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*. Boca Raton, FL: Chapman and Hall/CRC.

DEAN, C. AND LAWLESS, J. F. (1989). Comments on *An extension of quasi-likelihood estimation* by Godambe and Thompson. *Journal of Statistical Planning and Inference* **22**, 155–158.

DERSIMONIAN, R. AND LAIRD, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

GOURIEROUX, C., MONFORT, A. AND TROGNON, A. (1984). Pseudo maximum likelihood methods: theory. *Econometrica* **52**, 681–700.

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

HARDY, R. J. AND THOMPSON, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* **15**, 619–629.

LACHIN, J. M. (2000). *Biostatistical Methods. The Assessment of Relative Risks*. New York: Wiley.

LEE, Y. G. AND NELDER, J. A. (2000). Two ways of modelling overdispersion in non-normal data. *Applied Statistics* **49**, 591–598.

LAWSON, A., BIGGERI, A., BÖHNING, D., LESAFFRE, E., VIEL, J.-F. AND BERTOLLINI, R. (1999). *Disease Mapping and Risk Assessment for Public Health*. New York: Wiley.

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

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

MOORE, D. F. (1986). Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika* **73**, 583–588.

NELDER, J. A. AND LEE, Y. G. (1998). Joint modelling of mean and dispersion. *Technometrics* **40**, 168–171.

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.

WOODWARD, M. (1999). *Epidemiology. Study Design and Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

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

[*Received June 10, 2002; first revision December 12, 2002; second revision June 5, 2003; accepted for publication June 23, 2003*]