Citation:

Malzahn, U. (2003). Meta-analysis: A general principle for estimating heterogeneity variance in several models. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 41-52). Hogrefe & Huber.

3

Meta-Analysis: A General Principle for Estimating Heterogeneity Variance in Several Models

Uwe Malzahn

Working Group: Biometry and Epidemiology
Institute for International Health, Joint Center for Humanities and Health Sciences
Free University Berlin

Summary

A main question in meta-analysis is the comparability of studies in consideration. This relates and leads inevitably to the investigation of problems of heterogeneity. In this chapter, we deal with the one-dimensional case, represented by four examples, and propose a nonparametric moment estimator for the heterogeneity variance in the corresponding random effects model. The principle is based on decomposing the variance of the study estimator, that is, the total (unconditional) variance is composed of the mean conditional variance and the heterogeneity variance, or an expression containing the latter. We also hint to problems concerning the use of the DerSimonian-Laird estimator, which is a frequently used nonparametric estimator of general application. Finally, based on expressions for the conditional variances in several models for effect parameters and quality scores we demonstrate our principle.

3.1 INTRODUCTION AND EXAMPLES

We denote by $\theta \in \mathbb{R}$ the measure of the effect of interest. Given population heterogeneity, there is an increase of the variance for the study estimator. It appears an additional variance term, the heterogeneity variance τ^2. An estimation $\hat{\tau}^2$ for τ^2 can be used to adapt the inference regarding the overall mean of θ.

We suppose that the meta-analysis is based on k studies, or charges of a pharmaceutical product (solution, powder), respectively. Let us consider four examples for an effect measure or quality score θ.

Standardized difference This effect size measure is used for comparisons of groups based on continuous measurement variables with (possibly) different scales of measurement:

$$\theta = \frac{(\mu^T - \mu^C)}{\sigma_{T,C}}.$$

Here, μ^T and μ^C denote the mean values in a treatment and control group, $\sigma_{T,C}$ denotes the variance of the response variables in the two groups. That means equal variances of the two groups within each study are assumed.

Standardized mortality ratio Here, we are interested in the expected number of counts for a case event in a region or area in comparison to the corresponding number in a reference population with the same population structure:

$$\theta = \frac{\mu}{e}.$$

At this μ is the mean number of mortality or morbidity cases for a geographic region or area, and e is the corresponding value for this area calculated on the basis of an external reference population. Clearly, both values depend on the population size for the area considered.

Log relative risk in Cox regression with random censorship We consider the log-linear Cox-model with only one covariate. This covariate is a dichotomous variable which indicates some grouping membership (new therapy/treatment – standard therapy/placebo). Here, the parameter θ can be interpreted as the *logarithm of the relative risk* ($\ln(RR)$):

$$\theta = \ln(RR) = \ln\left(\frac{\lambda(t|Z=1)}{\lambda(t|Z=0)}\right) = \ln\left(\frac{\lambda^T(t)}{\lambda^C(t)}\right),$$

where $\lambda(\cdot|Z) = \lambda_0(\cdot)e^{\theta Z}$ denotes the hazard rate function. We suppose that the distribution function $F_0(t) = P(T \leq t|Z=0)$ is continuous.

$S_0(t) = \exp\left(-\int_0^t \lambda_0(z)dz\right)$ is the survival function corresponding to $\lambda_0(\cdot)$.

Quality scores (in pharmaceutical technology) This type of scores is used for in-process control detecting polluting particles in solutions and powders:

$$\theta = \frac{1}{n} c_0 \left(\lambda_l + c_m \lambda_m + c_g \lambda_g \right).$$

Here, n denotes the size of the sample which is drawn from each charge of the product; λ_l, λ_m, and λ_g denote the expected numbers of slight, moderate, and severe faults (pollutions, contaminations) in the sample; c_m and c_g are coefficients to weigh the kinds of fault.

The (unknown) study-/charge-specific values of the effect-/score parameter are denoted by θ_i; $\hat{\theta}_i$ is the estimate in the study number i. Homogeneity means that $\theta_1 = \theta_2 = \cdots = \theta_k$.

In the random effects model (RE model) we have to distinguish between the conditional distribution of the random variable $\hat{\theta}$, given a fixed study-specific parameter value θ, $P^{\hat{\theta}}$, and the distribution of the parameter θ in the population of study parameters. In the context of heterogeneity analysis the latter distribution is called the heterogeneity distribution, say G.

We will assume that $\hat{\theta}$ is a conditional unbiased estimator[1], and $\hat{\theta}$ is regarded as the bias corrected version of an estimator $\tilde{\theta}$. In the following, we give the study estimators for our four examples.

Standardized difference

$$\hat{\theta}_i = \left[H(N_i/2) \right]^{-1} \frac{\left(\overline{X}_i^T - \overline{X}_i^C \right)}{s_i^2}.$$

Here, s_i^2 denotes the pooled sample variance, and $H(N_i/2)$ is the bias correcting factor:

$$H(a) = \sqrt{a} \frac{\Gamma(a - \frac{1}{2})}{\Gamma(a)}, \text{ and furthermore, } N_i = n_i^T + n_i^C - 2,$$

in which n_i^T and n_i^C are the group sizes in study i, and $\Gamma(\cdot)$ denotes the gamma function.

Standardized mortality ratio

$$\hat{\theta}_i = \frac{Y_i}{e_i}.$$

Here, Y_i is the observed number of mortality or morbidity cases in region i, and e_i is the corresponding expected number calculated on the basis of an external reference population.

[1]More precisely, we only need that $E\left(\hat{\theta}|\theta\right) = \theta + C$, in which the constant C does not depend on θ.

Log relative risk in Cox regression with random censorship Here, $\hat{\theta}_i$ is the maximum partial likelihood estimator (MPLE), which is asymptotically unbiased:

$$\hat{\theta}_i = \text{argmax}_\theta \prod_{j=1}^{L_i} \frac{\exp(\theta_i Z_{i(j)})}{\sum_{l \in R_{ij}} \exp(\theta_i Z_{il})}$$

and n_i denotes the number of individuals at the beginning of study i, $T_{i1}^0 < \cdots < T_{iL_i}^0$ are the failure times, and $Z_{i(j)}$ is the covariate value for the individual failed at time T_{ij}^0. Furthermore, R_{ij} is the risk set immediately before time T_{ij}^0 in study i. The data are (X_{il}, δ_{il}), $X_{il} = \min(T_{il}, U_{il})$, in which T_{il} and U_{il} are the variables for the failure time and censoring time for individual l in study i, and $\delta_{il} = I_{\{T_{il} \leq U_{il}\}}$.

Quality scores

$$\hat{\theta}_i = \frac{c_0}{n_i} \left(l_i + c_m m_i + c_g g_i \right),$$

where l_i, m_i and g_i are the observed numbers of slight, moderate, and severe faults in charge i. Here, we usually have $n_i \equiv n$.

Let us summarize the assumptions of the RE model:

$$\hat{\theta}_i = \theta_i + \varepsilon_i, \quad \theta_i = \mu_G + \xi_i, \tag{3.1}$$

in which the ε_i are independent random variables with

$$E(\varepsilon_i) = 0, \quad \nu_i^2 = \text{Var}(\varepsilon_i) = E_G\left(\sigma_i^2(\theta_i)\right), \text{with } \sigma_i^2(\theta_i) = \text{Var}\left(\hat{\theta}_i | \theta_i\right),$$
$$\xi_i \text{ i.i.d., } E(\xi_i) = 0, \qquad \text{Var}(\xi_i) = \text{Var}_G(\xi) = \tau^2, \quad \mu_G = E_G(\theta_i).$$

The conditional variances possibly depend both on the study design in study i and on the parameter θ_i. We consider the problem of estimating the heterogeneity variance τ^2.

3.2 A VARIANCE DECOMPOSITION

In this section, we generally denote by $\hat{\theta}$ a study-/charge-estimator with

$$\mu(\theta) = E\left(\hat{\theta} | \theta\right),$$
$$\sigma^2(\theta) = \text{Var}\left(\hat{\theta} | \theta\right).$$

Note, that more precisely we have to denote $\sigma^2(\theta; \Xi; \alpha_1, \cdots, \alpha_p)$, in which Ξ stands for the study design or, more generally, for characteristics of the experiment, for instance $N = n^T + n^C - 2$ for the standardized difference. Under the assumption θ random, $\theta \sim G$, we can decompose the total (unconditional)

variance:

$$\begin{aligned}\text{Var}\left(\hat{\theta}\right) &= E_G\left(\text{Var}\left(\hat{\theta}|\theta\right)\right) + \text{Var}_G\left(E\left(\hat{\theta}|\theta\right)\right) \\ &= \int \sigma^2\left(\theta\right) g\left(\theta\right) d\theta + \int \left(\mu\left(\theta\right) - \mu_G\right)^2 g\left(\theta\right) d\theta,\end{aligned} \tag{3.2}$$

where $g(\cdot)$ is the density or the probability mass function (which gives the single probabilities) in the case of a discrete heterogeneity distribution.

In the case that $\hat{\theta}$ is conditionally unbiased: $\mu(\theta) = \theta$, or if $\mu(\theta) = \theta + \text{const.}$, it follows that $\text{Var}_G\left(E\left(\hat{\theta}|\theta\right)\right) = \text{Var}_G\left(\theta\right) = \tau^2$, and

$$\tau^2 = \text{Var}\left(\hat{\theta}\right) - E_G\left(\text{Var}\left(\hat{\theta}|\theta\right)\right) \tag{3.3}$$

(see Equation 3.2). Equation 3.3 will motivate a principle for estimating τ^2. The advantages of the method are:

- the resulting estimator is very easily calculated,
- we avoid any parametric assumption about G,
- using $\text{Var}\left(\hat{\theta}|\theta\right)$, it is possible to take the special statistical model into account, that is, the special estimating problem.

This method is applicable under the supposition that we can express

$$E_G\left(\text{Var}\left(\hat{\theta}|\theta\right)\right) = F\left(\Lambda; \mu_G^{(l)}; E_G\left(\alpha_s^r\right)\right), \tag{3.4}$$

in which Λ comprises known quantities from the study design. Examples for α_s are $\alpha_1 = p_l$, $\alpha_2 = p_m$, and $\alpha_3 = p_g$ in the case of quality scores. $\sigma^2(\theta) = \tilde{F}(\Lambda; \theta^l; \alpha_s^r)$ is sufficient for Equation 3.4.

3.3 THE DERSIMONIAN-LAIRD ESTIMATOR

A simple, general, and frequently used method to estimate τ^2 is the *DerSimonian-Laird estimator*. Generally, this estimator can be derived without normality assumptions by means of the weighted least squares principle. For this, it is assumed that the conditional variances are *known*, the so-called study specific variances. We write ν_i^2 instead of $\sigma_i^2(\theta_i)$ because it makes no sense to assume on the one side that θ_i is an unknown realization of a random variable, furthermore σ_i^2 is known exactly, and on the other side, that we have a structural dependence of σ_i^2 on $\theta_i : \sigma_i^2(\theta_i)$.

Now we can write the model (see Equations 3.1) in vector notation with the "design matrix" $X = \mathbf{1}_k = (1, \ldots, 1)^T$:

$$\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right)^T = D = X\beta + F = \mu_G \mathbf{1}_k + (E + C),$$

in which

$$\begin{aligned} E &= (\varepsilon_1, \ldots, \varepsilon_k)^T, \\ C &= (\xi_1, \ldots, \xi_k)^T, \\ E\,(E + C) &= \mathbf{0}_k, \\ W :&= \mathrm{Cov}\,(E + C) = \tau^2 \mathbf{I}_k + V. \end{aligned}$$

Here, $\mathbf{I}_k$ denotes the k-dimensional identity matrix and $V = \mathrm{diag}(\nu_1^2, \ldots, \nu_k^2)$. Then it is straightforward to derive the weighted least squares (WLS) estimate $X\hat{\beta}_{\mathrm{WLS}} = \hat{\mu}_G$ with weighting matrix V^{-1} (note, that τ^2 is *unknown* but the ν_i^2 are assumed to be *known*). We have

$$\hat{\mu}_G = \left(\sum_{i=1}^{k} \nu_i^{-2} \hat{\theta}_i \right) / \left(\sum_{i=1}^{k} \nu_i^{-2} \right).$$

The corresponding sum of squared residuals is

$$RSS = |D - \hat{D}|^2_{V^{-1}} = \sum_{i=1}^{k} \nu_i^{-2} \hat{\theta}_i^2 - \left(\left(\sum_{i=1}^{k} \nu_i^{-2} \hat{\theta}_i \right)^2 / \left(\sum_{i=1}^{k} \nu_i^{-2} \right) \right),$$

with

$$E\,(RSS) = (k-1) + \left(\sum_{i=1}^{k} \nu_i^{-2} - \left(\sum_{i=1}^{k} \nu_i^{-4} / \sum_{i=1}^{k} \nu_i^{-2} \right) \right) \tau^2.$$

Rearranging this equation and replacing the expectation $E\,(RSS)$ by its observed value RSS, it follows

$$\hat{\tau}^2_{dl} = \frac{(RSS - (k-1))}{(S_1 - (S_2/S_1))} = \frac{\left(\sum_{i=1}^{k} \nu_i^{-2} \left(\hat{\theta}_i - \hat{\mu}_G \right)^2 - (k-1) \right)}{(S_1 - (S_2/S_1))}, \tag{3.5}$$

with $S_l = \sum_{i=1}^{k} \left(\nu_i^{-2} \right)^l$, $l = 1, 2$.

However, for the application in practice the true study specific variances ν_i^2 are *unknown*, that means, for the data analysis they are *estimated*. Additionally, in most applications we have $\mathrm{Var}\left(\hat{\theta}|\theta\right) = \sigma^2(\theta)$ and $\sigma^2\left(\alpha_1, \ldots, \alpha_p\right)$, respectively, with unknown θ and α_s. Note, that in Equation 3.5 we have $\hat{\mu}_G = \hat{\mu}_G(\nu_1^2, \ldots, \nu_k^2)$ and $S_l = S_l(\nu_1^2, \ldots, \nu_k^2)$.

The problem is: What do we have to put in for $(\nu_1^2, \ldots, \nu_k^2)$? It would be good practice to start from an adequate model, find the right conditional distribution in this model, and with this derive the expression for the conditional variances $\sigma_i^2(\theta_i)$. Finally, we put in

$$\nu_i^2 := \hat{\sigma}_i^2(\theta_i) = \sigma_i^2\left(\hat{\theta}_i\right),$$

to obtain a practical version of the DerSimonian-Laird estimator. Note, that $\hat{\theta}_i$ is random. Consequently, $\hat{\tau}^2_{dl}$ is no longer the best linear unbiased estimator because the optimal weights v_i^{-2} are unknown. Moreover, τ^2_{dl} is not unbiased, numerator *and denominator* in Equation 3.5 are stochastic terms.

3.4 THE CONDITIONAL VARIANCES IN THE MODELS

For the Examples

Standardized difference Under the assumption of normally distributed measurement variables

$$X_{ij} \sim \mathcal{N}(\mu_i^T, \sigma^2_{i;T,C}),\ j = 1, \ldots, n_i^T,$$
$$Y_{ij} \sim \mathcal{N}(\mu_i^C, \sigma^2_{i;T,C}),\ j = 1, \ldots, n_i^C$$

it follows that

$$\sqrt{q_i} H\left(N_i/2\right) \hat{\theta}_i \sim t_{N_i}\left(\theta_i \sqrt{q_i}\right),$$

a noncentral t-distribution with N_i degrees of freedom and noncentrality parameter $\theta_i \sqrt{q_i}$, in which $q_i = n_i^T n_i^C / \left(n_i^T + n_i^C\right)$. Therefore, we have

$$\sigma_i^2\left(\theta_i\right) = \left(H\left(N_i/2\right)\right)^{-2} \frac{N_i}{q_i\left(N_i - 2\right)} + \left(\frac{N_i}{\left(N_i - 2\right)} \left(H\left(H_i/2\right)\right)^{-2} - 1\right) \theta_i^2 \quad (3.6)$$

(see Malzahn, Böhning, & Holling, 2000).

Standardized mortality ratio Since conditional on the value θ_i in area i, a Poisson distribution with parameter $\lambda_i = \theta_i e_i$ is assumed for Y_i, it is easy to see that

$$\sigma_i^2 = \frac{\theta_i}{e_i}.$$

Maximum partial likelihood estimator for survival time studies It can be shown (see Fleming & Harrington, 1994) that

$$n_i^{1/2}\left(\hat{\theta}_i^{(n_i)} - \theta_i\right) \longrightarrow_L X,$$

with $X \sim \mathcal{N}\left(0, \sigma_i^{-2}\left(\theta_i\right)\right)$, in which the inverse of the asymptotical variance for the standardized estimator is under additional assumptions (in order to

reduce the complexity of the resulting expression):

$$\begin{aligned}\sigma_i^2(\theta_i) &= \frac{1}{2}\exp(\theta_i)\int_0^{t_{up}}\left(1-\frac{t}{u_i}\right)\frac{(S_0(t))^{\exp(\theta_i)}}{\left[\exp(\theta_i)(S_0(t))^{\exp(\theta_i)}+S_0(t)\right]}f_0(t)\,dt \\ &= \frac{1}{2}\exp(\theta_i)\int_{S_0(t_{up})}^{1}\left(1-\frac{S_0^{-1}(s)}{u_i}\right)\frac{s^{\exp(\theta_i)}}{\left[\exp(\theta_i)s^{\exp(\theta_i)}+s\right]}ds.\end{aligned} \tag{3.7}$$

Here, t_{up} denotes a constant, common for all studies (it depends on the baseline hazard rate function, which is taken as a basis), and u_i are (possibly study-specific) constants, characterizing the censoring time distribution; S_0 denotes the survival function according to the baseline hazard: $S_0(s) = P_0(T > s)$.

Quality scores The situation can be described by a multinomial distribution model $M(n; p_l, p_m, p_g)$ with probability mass function:

$$\pi(l,m,g|\mathbf{p}) = \frac{n!}{l!m!g!}p_l^l p_m^m p_g^g (1-p_l-p_m-p_g)^{n-l-m-g}. \tag{3.8}$$

Here, $\mathbf{p} = (p_l, p_m, p_g)^T$ denotes the vector of the probabilities for detecting a slight, moderate, or severe contamination in an inspected item. For heterogeneity analysis it is important that we interpret the expression in Equation 3.8 as a *conditional* distribution: the distribution for the vector $(l, m, g)^T$ at fixed underlying vector $(p_l, p_m, p_g)^T$. Heterogeneity means: There exists a non-degenerated heterogeneity distribution G on $(0,1)^3$, and for each charge of the product under examination the actually underlying parameter vector $\mathbf{p}$ is a realization from this distribution. In this model, the conditional[2] variance of the quality score in charge i is

$$\begin{aligned}\sigma^2\left(\mathbf{p}^{(i)}\right) &:= \mathrm{Var}\left(\hat{\theta}_i|\mathbf{p}^{(i)}\right) \\ &= \frac{c_0^2}{n}\Big[p_l^{(i)}\left(1-p_l^{(i)}\right)+c_m^2 p_m^{(i)}\left(1-p_m^{(i)}\right)+c_g^2 p_g^{(i)}\left(1-p_g^{(i)}\right) \\ &\quad -2c_m p_l^{(i)}p_m^{(i)}-2c_g p_l^{(i)}p_g^{(i)}-2c_m c_g p_m^{(i)}p_g^{(i)}\Big].\end{aligned}$$

3.5 A PRINCIPLE TO ESTIMATE THE HETEROGENEITY VARIANCE

Our starting point here is the relationship given in Equation 3.3 for the heterogeneity variance. If θ^2 enters in the expression for $\mathrm{Var}\left(\hat{\theta}|\theta\right)$, then τ^2 enters in

[2]Conditional on fixed $\mathbf{p}^{(i)} = \left(p_l^{(i)}, p_m^{(i)}, p_g^{(i)}\right)^T$.

an analogous manner in the right side of Equation 3.3, leading to an equation for τ^2 which is specifically considered for the model. This equation provides the possibility to construct an estimator $\hat{\tau}^2$. As an example, we want to demonstrate this principle for the standardized difference (see Malzahn et al., 2000).

Standardized difference Here, Equation 3.6 together with the relationship $E_G\left(\theta^2\right) = \tau^2 + \mu_G^2$ yields

$$\int \sigma^2(\theta)\, g(\theta)\, d\theta = \left(H\left(N/2\right)\right)^{-2} \frac{N}{q\left(N-2\right)} + \left[\frac{N}{\left(N-2\right)}\left(H\left(N/2\right)\right)^{-2} - 1\right]\left(\tau^2 + \mu_G^2\right).$$

Applying this and rearranging leads to

$$\tau^2 = \left(H\left(N/2\right)\right)^2 \frac{\left(N-2\right)}{N} \operatorname{Var}\left(\hat{\theta}\right) - q^{-1} - \left[1 - \frac{\left(N-2\right)}{N}\left(H\left(N/2\right)\right)^2\right]\mu_G, \tag{3.9}$$

where $N = n^T + n^C - 2$ and $q = n^T n^C / (n^T + n^C)$.

The data are $\left(\hat{\theta}_1; N_1, q_1\right), \ldots, \left(\hat{\theta}_k; N_k, q_k\right)$. Equation 3.9 will motivate a nonparametric estimator $\hat{\tau}^2$. To estimate the first term of Equation 3.9, it seems to be reasonable to use a modified version of the usual empirical variance of the study estimators, considering the different degrees of freedom (N_i). We can estimate the mean value of the effect parameter in the overall population by

$$\hat{\mu}_{\hat{\theta}} = k^{-1} \sum_{i=1}^{k} \hat{\theta}_i$$

or, given estimates $\hat{v}_i^2$ of the study-specific variances for $\hat{\theta}_i$, the pooled estimator

$$\hat{\mu}_{\hat{\theta}} = \frac{\sum_{i=1}^{k} \hat{v}_i^{-2} \hat{\theta}_i}{\sum_{i=1}^{k} \hat{v}_i^{-2}}.$$

In the fixed effects model for known study-specific variances, the pooled mean is the best unbiased linear estimator for the first moment and should be used if the data indicates at most a small heterogeneity variance. In the case of large heterogeneity, the arithmetic mean should be preferred because the "true" weights within the pooled estimator are poorly estimated by noniterative procedures. Finally, in Equation 3.9 we estimate

$$\left[1 - \frac{\left(N-2\right)}{N}\left(H\left(N/2\right)\right)^2\right]\mu_G^2$$

by the mean value of the corresponding study specific realizations. This leads to a nonparametric estimator of the heterogeneity variance given by

$$\hat{\tau}^2 = \frac{1}{(k-1)} \sum_{i=1}^{k} (1 - K_i) \left(\hat{\theta}_i - \hat{\mu}_{\hat{\theta}}\right)^2 - \frac{1}{k} \sum_{i=1}^{k} \frac{1}{q_i} - \frac{1}{k} \sum_{i=1}^{k} K_i \hat{\theta}_i^2,$$

where

$$K_i = 1 - \left(H\left(N_i/2\right)\right)^2 \frac{(N_i - 2)}{N_i}.$$

The same mode of procedure yields corresponding estimators for the heterogeneity variance in the case of the standardized mortality ratio.

Standardized mortality ratio

$$\hat{\tau}^2 = \frac{1}{(k-1)} \sum_{i=1}^{k} \left(\hat{\theta}_i - \hat{\mu}_{\hat{\theta}}\right)^2 - \frac{1}{k} \hat{\mu}_{\hat{\theta}} \sum_{i=1}^{k} \frac{1}{e_i},$$

where $\hat{\mu}_{\hat{\theta}}$ is an estimator for the mean value of the parameter in the whole population. Typically, two estimators are considered:

the arithmetic mean

$$\hat{\mu}_{\hat{\theta}}^{(ar)} = \frac{1}{k} \sum_{i=1}^{k} \hat{\theta}_i$$

and the pooled mean

$$\hat{\mu}_{\hat{\theta}}^{(pool)} = \frac{\sum_{i=1}^{k} \hat{\theta}_i e_i}{\sum_{i=1}^{k} e_i}$$

(see Böhning, Sarol, & Malzahn, 2000).

Log relative risk in the log-linear Cox model with random censorship At first, we consider the inverse of the asymptotical conditional variance of the standardized estimator in this model, given by Equation 3.7. This quantity has to be estimated for each study. The expression in Equation 3.7 for $\sigma_i^2(\theta_i)$ contains the unknown survival function $S_0(t)$ corresponding to the baseline hazard. An obvious nonparametric estimator $\hat{S}_0(t)$ in study i is the Kaplan-Meier estimator

$$\hat{S}_0(t) = \prod_{j:T_{ij}^0 \leq t} \frac{\left(\bar{Y}_{ij} - 1\right)}{\bar{Y}_{ij}},$$

in which

$$\bar{Y}_{ij} = \sum_{l=1}^{n_i} Y_l \left(T_{ij}^0\right), \quad \text{with} \quad Y_l(t) = I_{(X_l \geq t)}.$$

Furthermore, since there are no bindings, the Nelson estimator for the cumulative hazard function

$$\Lambda_0(t) = \int_0^t \lambda_0(z)dz$$

is given by

$$\hat{\Lambda}_0 = \sum_{j:T_{ij}^0 \leq t} \frac{\delta_{ij}}{\overline{Y}_{ij}}.$$

Consequently, a natural estimator for an integral of the form $\int_a^b h_i(t)\lambda_0(t)dt$ is given by

$$\sum_{j:a<T_j^0<b} h_i\left(T_j^0\right)\frac{\delta_j}{\overline{Y}_j}.$$

Because of $f_0(s) = \lambda_0(s)S_0(s)$, we have

$$h_i(t) = \frac{\left(S_0\left(t\right)\right)^{\exp(\theta_i)+1}}{\left[\exp\left(\theta\right)\left(S_0(t)\right)^{\exp(\theta_i)} + S_0(t)\right]}\left(1 - \frac{1}{u_i}\right)$$

in Equation 3.7, leading to the estimator

$$\begin{aligned}\hat{\sigma}_i^2(\theta_i) =& \frac{\exp(\hat{\theta}_i)}{2} \sum_{j:T_{ij}^0 \leq t_{up}} \frac{\left(\hat{S}_0\left(T_{ij}^0\right)\right)^{\exp(\hat{\theta}_i)+1}}{\left[\exp(\hat{\theta}_i)\left(\hat{S}_0\left(T_{ij}^0\right)\right)^{\exp(\hat{\theta}_i)} + \hat{S}_0\left(T_{ij}^0\right)\right]} \\ &\times \left(1 - \frac{T_{ij}^0}{u_i}\right)\frac{\delta_{ij}}{\overline{Y}_{ij}},\end{aligned}$$

where $\overline{Y}_{ij}$ is the size of the risk set at failure time T_{ij}^0, and $\delta_{ij} := I_{\left(T_{ij} \leq U_{ij}\right)}$, that is, $\delta_{ij} = 1$ if the individual number j in study i is a failure, and $\delta_{ij} = 0$ if this individual is a censored observation. Because the MPLE $\hat{\theta}$ is asymptotically unbiased, Equation 3.3 suggests estimators of the form

$$\hat{\tau}^2 = \frac{1}{(k-1)}\sum_{i=1}^{k}\left(\hat{\theta}_i - \hat{\mu}_{\hat{\theta}}\right)^2 - \hat{E}_G\left(\left(n\hat{\sigma}^2\left(\hat{\theta}\right)\right)^{-1}\right),$$

where

$$\hat{E}_G\left(\left(n\hat{\sigma}^2\left(\hat{\theta}\right)\right)^{-1}\right) = H\left(n_1^{-1}\hat{\sigma}_1^{-2}\left(\hat{\theta}_1\right), \ldots, n_k^{-1}\hat{\sigma}_k^{-2}\left(\hat{\theta}_k\right)\right).$$

The most simple case is $H(x_1, \ldots, x_k) = \overline{x}$, but this does not make much sense, rather it seems to be more sensible to derive a weighted mean of the

$n_i^{-1}\hat{\sigma}_i^{-2}(\hat{\theta}_i)$. Currently, this is an open problem and will be subject of further research.

Quality scores Here we can derive:

$$\begin{aligned}\hat{\tau}^2 = & \frac{1}{(k-1)}\sum_{i=1}^{k}\left(\hat{\theta}_i - \hat{\mu}_{\hat{\theta}}\right)^2 \\ & - \frac{c_0^2}{n(n-1)}\left[\overline{b} + c_m^2\overline{m} + c_g^2\overline{g}\right] + \frac{c_0^2}{n(n-1)}\left[\overline{l^2} + c_m^2\overline{m^2} + c_g^2\overline{g^2}\right] \\ & - 2\frac{c_0^2}{n^2}\left[c_m\frac{1}{(k-1)}\sum_{i=1}^{k}\left(l_i - \overline{l}\right)(m_i - \overline{m}) + c_g\frac{1}{(k-1)}\sum_{i=1}^{k}\left(l_i - \overline{l}\right)(g_i - \overline{g})\right. \\ & \left. + c_m c_g\frac{1}{(k-1)}\sum_{i=1}^{k}(m_i - \overline{m})(g_i - \overline{g})\right],\end{aligned}$$

where

$$\overline{l^\alpha} = \frac{1}{k}\sum_{i=1}^{k} l_i^\alpha, \quad \overline{l} = \overline{l^1}.$$

REFERENCES

Böhning, D., Sarol, J., & Malzahn, U. (2000). *Efficient non-iterative and nonparametric estimation of heterogeneity variance for the standardised mortality ratio.* (unpublished manuscript).

Fleming, T. R., & Harrington, D. P. (1994). *Counting processes and survival analysis.* New York: Wiley.

Malzahn, U., Böhning, D., & Holling, H. (2000). Nonparametric estimation of heterogeneity variance for the standardised difference used in meta–analysis. *Biometrika*, *87*, 619–632.