SIM 2710
pp: 1–20 (col.fig.: nil)
PROD. TYPE: COM
ED: Durga
PAGN: Vijay – SCAN: Naveen

STATISTICS IN MEDICINE
*Statist. Med.* 2006; **25**:000–000
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2710



# A comparison of three different models for estimating relative risk in meta-analysis of clinical trials under unobserved heterogeneity

Ronny Kuhnert[1,*,†] and Dankmar Böhning[2]

[1]*Division for International Health and Biometry, Institute for Social Medicine, Epidemiology, and Health Economy, Joint Centre for Health Sciences and Humanities, Charité Medical School Berlin, Fabeckstr. 60-62, 14195 Berlin, Germany*
[2]*Applied Statistics, School of Biological Sciences, The University of Reading Harry Pitt Building, Earley Gate, Whiteknights Road, Reading RG6 6FN, U.K.*

SUMMARY

We focus on the comparison of three statistical models used to estimate the treatment effect in meta-analysis when individually pooled data are available. The models are two conventional models, namely a multi-level and a model based upon an approximate likelihood, and a newly developed model, the *profile likelihood* model which might be viewed as an extension of the Mantel–Haenszel approach. To exemplify these methods, we use results from a meta-analysis of 22 trials to prevent respiratory tract infections. We show that by using the multi-level approach, in the case of baseline heterogeneity, the number of clusters or components is considerably over-estimated. The approximate and profile likelihood method showed nearly the same pattern for the treatment effect distribution. To provide more evidence two simulation studies are accomplished. The profile likelihood can be considered as a clear alternative to the approximate likelihood model. In the case of strong baseline heterogeneity, the profile likelihood method shows superior behaviour when compared with the multi-level model. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: MAIPD; multi-level model; profile likelihood; approximate likelihood; mixture model; unobserved heterogeneity



*Correspondence to: Ronny Kuhnert, Robert-Koch Institute, Division for Health of Children and Adolescents, Prevention Concepts, Seestr. 10, 13353 Berlin, Germany.
†E-mail: kuhnertr@rki.de

*Received 20 October 2005*
*Accepted 2 August 2006*

## 1. INTRODUCTION

This contribution focuses on the comparison of three models used to estimate treatment effect in meta-analysis under unobserved heterogeneity. The models compared consist of two conventional models, namely the multi-level (ML) and the approximate likelihood (AL) model, and a newly developed model for meta-analysis with individually pooled data, namely the profile likelihood (PL) model. In recent years, meta-analysis has become an essential method used to provide more profound information on the effect of a treatment or intervention and has been demonstrated to provide a powerful statistical tool to combine and analyse the results from individual studies. Numerous international publications prove the quality and the common practicability of meta-analysis. An introduction and an explanation on this issue are given by Olkin [1] and Jones [2] in this journal or by Fleiss [3]. Important for our situation here is the availability of the number of events $x_i^{\mathrm{T}}$ ($x_i^{\mathrm{C}}$) and the number under risk $n_i^{\mathrm{T}}$ ($n_i^{\mathrm{C}}$) in the treatment arm (control arm) for each trial $i$ involved in the meta-analysis of a total of $k$ studies. We call this situation of meta-analysis a *meta-analysis using individually pooled data* (MAIPD). Note that frequently a MAIPD could be done just on the basis of the published literature (for a collection of numerous meta-analysis of this type see Reference [4]). However, individual patient data are typically *not* available from a MAIPD and are difficult, not to say usually impossible to retrieve from a MAIPD. Note further that in contrast to conventional meta-analysis where typically an effect measure accompanied by a variance estimate is available (see for example Reference [5]), a MAIPD offers more choices for the analyst, for example in choosing different effect measures such as relative risk, risk difference, number-needed-to-treat with accompanying variances. In this contribution we will exploit another option of MAIPD, namely we will use the *profile likelihood* method to eliminate the nuisance baseline parameter. Consecutively, the inference will be based on the profile likelihood. In particular, the method is able to cope with the two forms of heterogeneity that can occur in a MAIPD: the *baseline heterogeneity* that arises in the control arm and the *effect heterogeneity* that arises from potential different treatment effects in the various trials.

Another issue is the connection of MAIPD to the analysis of multicentre studies, frequently mentioned in the literature. Note that in a multicentre clinical trial the methodology used in MAIPD is validly applicable as well. However, multicentre study data are often neither published completely, nor are they made available for numerous reasons to the interested reader, although centre specific information might be found in the corresponding publication. Therefore, methodology of MAIPD might also be very useful in this situation. Consequently, MAIPD is a common situation in multicentre clinical trials; for further discussion of this issue see Reference [6]. All three models discussed here are also applicable in multicentre studies. Bearing this in mind, one should see the clear differences between MAIPD in general on the one hand, and a multicentre trial on the other hand. For example, one difference between the two is the stronger guidelines in the case of multicentre studies. Therefore, the baseline heterogeneity in a MAIPD can be expected to be larger.

Let us turn now to an example of a typical framework of a MAIPD as shown in Table I. For this meta-analysis, 22 trials were included to investigate the effect of selective decontamination of the digestive tract on the risk of respiratory tract infection (see Reference [7]). Patients in intensive care units were randomly selected either to receive treatment through a combination of non-absorbable antibiotics, or no treatment. As already mentioned above, in a MAIPD two forms of heterogeneity might occur. The first form is the baseline heterogeneity, which can be found when the incidence varies in the control arm. Baseline heterogeneity is not the focus of a MAIPD





DOI: 10.1002/sim

Table I. Respiratory tract infections in treated and control groups of 22 centres.

| Study | Treatment | | Control | |
|---|---|---|---|---|
| $i$ | $x_i^T$ | $n_i^T$ | $x_i^C$ | $n_i^C$ |
| 1 | 7 | 47 | 25 | 54 |
| 2 | 4 | 38 | 24 | 41 |
| 3 | 20 | 96 | 37 | 95 |
| 4 | 1 | 14 | 11 | 17 |
| 5 | 10 | 48 | 26 | 49 |
| 6 | 2 | 101 | 13 | 84 |
| 7 | 12 | 161 | 38 | 170 |
| 8 | 1 | 28 | 29 | 60 |
| 9 | 1 | 19 | 9 | 20 |
| 10 | 22 | 49 | 44 | 47 |
| 11 | 25 | 162 | 30 | 160 |
| 12 | 31 | 200 | 40 | 185 |
| 13 | 9 | 39 | 10 | 41 |
| 14 | 22 | 193 | 40 | 185 |
| 15 | 0 | 45 | 4 | 46 |
| 16 | 31 | 131 | 60 | 140 |
| 17 | 4 | 75 | 12 | 75 |
| 18 | 31 | 220 | 42 | 225 |
| 19 | 7 | 55 | 26 | 57 |
| 20 | 3 | 91 | 17 | 92 |
| 21 | 14 | 25 | 23 | 23 |
| 22 | 3 | 65 | 6 | 68 |

but it is important for achieving valid inference. Each model discussed here has a different way of modelling baseline heterogeneity. The second form is the effect heterogeneity. In this contribution the effect is expressed by means of the relative risk $\theta$, defined as

$$\theta_i = \frac{p_i^T}{p_i^C} \quad \text{and estimated as} \quad \frac{x_i^T n_i^C}{x_i^C n_i^T} \tag{1}$$

and we will focus here on estimating heterogeneity of effect.

Effect heterogeneity can also be visualized graphically. In Figure 1, all studies are shown which are included in the meta-analysis of the effect of selective decontamination of the digestive tract on the risk of respiratory tract infection. Each study (each point in Figure 1) is characterized by the rate of infection in the control arm, $p_i^C$ (the horizontal axis in Figure 1), and the rate of infection in the treatment arm, $p_i^T$ (vertical axis in Figure 1), respectively. The dashed line is a reference line with slope 1, so that studies lying on this line would have identical infection rates, or $\theta_i = 1$. In the case of effect homogeneity, all trials would experience the same relative risk, and the associated lines connecting each study point with the origin would be close together. In more generality, the slope $\theta_i$ of any line through the origin in the diagram would correspond to the associated relative risk of study $i$. In the most general form of heterogeneity each study represents its own cluster or component. Consequently, each study has a unique line through the point of origin. As an example, we draw a line to the 10th study. Several studies, like the 15th and 17th, are very close to this





DOI: 10.1002/sim

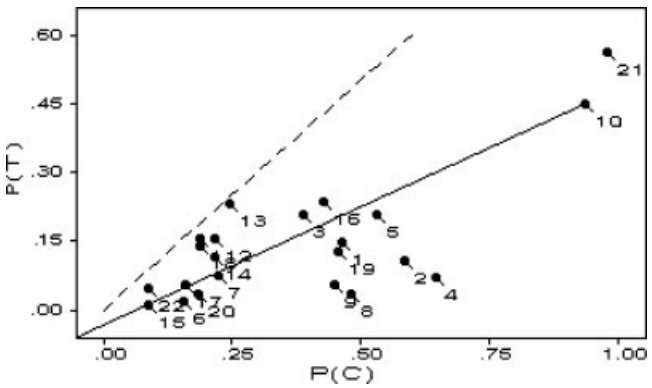


Figure 1. Representation of treatment effect as pairs of infection rate in control and treatment arm for the 22 trials of respiratory tract infection.

line. Consequently these studies have nearly the same relative risk, so that it can be assumed that several homogeneous clusters exist.

## 2. METHODS

We now turn to a more formal likelihood based modelling of heterogeneity in relative risk.

### 2.1. Modelling of heterogeneity

One possibility for modelling the cluster structure in the MAIPD is provided by means of the non-parametric mixture distribution as

$$f(x_i|P) = \sum_{j=1}^{m} f(x_i|\theta_j) q_j$$

where $f(x_i|\theta_j)$ is a parametric density, called the *mixture kernel*, $x_i = (x_i^{\mathrm{T}}, n_i^{\mathrm{T}}, x_i^{\mathrm{C}}, n_i^{\mathrm{C}})'$ contain the data for trial $i$,

$$P = \begin{pmatrix} \theta_1 & \dots & \theta_m \\ q_1 & \dots & q_m \end{pmatrix}$$

is a discrete mixing distribution with $q_j > 0$ and $\sum_{j=1}^{m} q_j = 1$, and where $m$ is the number of components or clusters.

The mixture model arises in a natural way as the marginal model with respect to an unobserved indicator variable $z_{ij}$ indicating if trial $i$ belongs to cluster $j$. Note that different forms of heterogeneity can be reproduced by the mixture distribution. If $m = 1$, then this is the homogeneity case, or if $m = k$ the number of studies, then this situation is the most general form of heterogeneity. Typically, the estimated number of clusters or components turns out to be low in comparison to the number of trials involved in the MAIPD. The cluster structure is mapped by means of the mixing distribution $P$.



DOI: 10.1002/sim

The *mixture kernel* is dependent on the model, respectively, and different kernels lead to the three different models under comparison, as will be briefly introduced in the next sections.

### 2.2. Approximate likelihood model

This approach considers the logarithmic relative risk $\phi_i = \log(\theta_i)$, which can be estimated as

$$\hat{\phi}_i = \log\left(\frac{x_i^{\mathrm{T}}}{n_i^{\mathrm{T}}}\right) - \log\left(\frac{x_i^{\mathrm{C}}}{n_i^{\mathrm{C}}}\right)$$

assuming non-zero events in both arms of trial $i$. Frequently, it is assumed that—conditional upon study $i$—the distribution of $\hat{\phi}_i$ might be validly approximated by a normal distributed with unknown mean $\phi_i$ and known standard deviation $\sigma_i$, though it is sufficient to assume that the log-likelihood is well approximated by a quadratic (see Reference [5]).

Therefore, the associated kernel $f(x_i|\theta_j)$ in the mixture model is the normal density, written as

$$f_{\mathrm{AL}}(\hat{\phi}_i|\theta_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \mathrm{e}^{-(\hat{\phi}_i - \log(\theta_i))^2/2\sigma_i^2} \tag{2}$$

Note that $\hat{\phi}_i$ is the log relative risk, a scalar, and the individually pooled counts for trial $i$ are not used. In the most general case, the variance is unknown and must be estimated. Under the assumption of Poisson distributed observations, the variance of the log relative risk in the $i$th study can be estimated using the delta-method (see Reference [8]) as

$$\widehat{\mathrm{var}}(\hat{\phi}_i) = \hat{\sigma}_i^2 = \frac{1}{x_i^{\mathrm{T}}} + \frac{1}{x_i^{\mathrm{C}}} \tag{3}$$

so that $\hat{\sigma}_i^2$ is used as a known parameter in (2). However, this simple formula has a number of drawbacks. Firstly, (3) is not defined if any of the trials has 0 successes. Secondly, for small trials, the normal approximation might not be satisfactory. Thirdly, the variance approximation used in the delta-method might be too crude to give a good approximation of the true variance of the log relative risk, especially if the event rate is different from 0.5. Nevertheless, the approach is quite popular (see References [9, 10])—likely due to its simplicity and lack of alternatives. To capture heterogeneity, the density (2) is used as kernel in the mixture distribution leading to

$$f(x_i|P) = \sum_{j=1}^{m} f_{\mathrm{AL}}(\hat{\phi}_i|\theta_j, \hat{\sigma}_i^2) q_j \tag{4}$$

In the homogeneity case the AL is frequently mentioned in the literature as the *fixed effect* model, see Reference [11]. The AL model was originally named fixed effects model by DerSimonian and Laird [12] for the case of homogeneity to distinguish it from the *random effects model* suggested by DerSimonian and Laird [12] to adjust for unobserved heterogeneity. The latter can be viewed as being further modelled with the above discrete mixture following Laird [13]. Here, we prefer to use the term AL model focussing on the fact the normal kernel in the mixture can be viewed as an approximating likelihood.





DOI: 10.1002/sim

### 2.3. Multi-level model

The ML model has become very popular in the biometrical literature (see Reference [14]) and captures the hierarchical structure of the data used in the meta-analysis. The first level (within study) can be modelled by means of the log-linear regression, with

$$\log(p_i^{\mathrm{C}}) = \alpha_i$$

$$\log(p_i^{\mathrm{T}}) = \alpha_i + \beta_i$$

$\alpha_i$ is in this case the baseline parameter and $\beta_i = \phi_i = \log(p_i^{\mathrm{T}}/p_i^{\mathrm{C}})$ is the effect parameter, the log-relative risk.[‡] Under the assumption that the observations are Poisson distributed, the likelihood for trial $i$ is given as

$$f_{\mathrm{ML}}(x_i|p_i^{\mathrm{C}}, p_i^{\mathrm{T}}) = \mathrm{e}^{-n_i^{\mathrm{T}} p_i^{\mathrm{T}}} \frac{(n_i^{\mathrm{T}} p_i^{\mathrm{T}})^{x_i^{\mathrm{T}}}}{x_i^{\mathrm{T}}!} \times \mathrm{e}^{-n_i^{\mathrm{C}} p_i^{\mathrm{C}}} \frac{(n_i^{\mathrm{C}} p_i^{\mathrm{C}})^{x_i^{\mathrm{C}}}}{x_i^{\mathrm{C}}!} \tag{5}$$

where $x_i$ is the vector $(x_i^{\mathrm{T}}, n_i^{\mathrm{T}}, x_i^{\mathrm{C}}, n_i^{\mathrm{C}})'$. The rates $p_i^{\mathrm{C}}$ and $p_i^{\mathrm{T}}$ can be replaced by their model associated parameters, namely

$$p_i^{\mathrm{C}} = \mathrm{e}^{\alpha_i}$$

$$p_i^{\mathrm{T}} = \mathrm{e}^{\alpha_i + \beta_i}$$

This leads to the following likelihood in trial $i$

$$f_{\mathrm{ML}}(x_i|\alpha_i, \beta_i) = \mathrm{e}^{-n_i^{\mathrm{T}} \mathrm{e}^{\alpha_i+\beta_i}} \frac{(n_i^{\mathrm{T}} \mathrm{e}^{\alpha_i+\beta_i})^{x_i^{\mathrm{T}}}}{x_i^{\mathrm{T}}!} \times \mathrm{e}^{-n_i^{\mathrm{C}} \mathrm{e}^{\alpha_i}} \frac{(n_i^{\mathrm{C}} \mathrm{e}^{\alpha_i})^{x_i^{\mathrm{C}}}}{x_i^{\mathrm{C}}!} \tag{6}$$

The second level is modelled by means of a non-parametric mixture distribution—as has been discussed and done previously. The most complex form of heterogeneity is considered, allowing baseline and effect heterogeneity, i.e. each mixed component has its own baseline and effect parameter. The mixture distribution has the form:

$$f(x_i|P) = \sum_{j=1}^{m} f_{\mathrm{ML}}(x_i|\alpha_j, \beta_j) q_j$$

$$\text{with } P = \begin{pmatrix} \alpha_1 & \dots & \alpha_m \\ \beta_1 & \dots & \beta_m \\ q_1 & \dots & q_m \end{pmatrix} \tag{7}$$

[‡]We use for this model the notation $\beta$ to keep the similarity with the existing literature.





DOI: 10.1002/sim

### 2.4. Profile-likelihood model

The PL approach has been well known for several decades (see References [15, 16]). In many cases, the PL method lacks practicability because the likelihood is rather complicated, difficult to handle, and often the profile likelihood itself exists only in iterated form. Nevertheless, the PL model is considered here since the PL-function is simple and useful in this special situation. Before coming to this result, let us consider the Poisson density for each centre

$$f_{\mathrm{PL}}(x_i|p_i^{\mathrm{C}}, p_i^{\mathrm{T}}) = \mathrm{e}^{-n_i^{\mathrm{T}} p_i^{\mathrm{T}}} \frac{(n_i^{\mathrm{T}} p_i^{\mathrm{T}})^{x_i^{\mathrm{T}}}}{x_i^{\mathrm{T}}!} \times \mathrm{e}^{-n_i^{\mathrm{C}} p_i^{\mathrm{C}}} \frac{(n_i^{\mathrm{C}} p_i^{\mathrm{C}})^{x_i^{\mathrm{C}}}}{x_i^{\mathrm{C}}!} \tag{8}$$

Taking the logarithms and ignoring terms not involving $p_i^{\mathrm{T}}$ or $p_i^{\mathrm{C}}$ in (8) we achieve that

$$lf_{\mathrm{PL}}(x_i|p_i^{\mathrm{T}}, p_i^{\mathrm{C}}) = x_i^{\mathrm{T}} \log(p_i^{\mathrm{T}}) - n_i^{\mathrm{T}} p_i^{\mathrm{T}} + x_i^{\mathrm{C}} \log(p_i^{\mathrm{C}}) - n_i^{\mathrm{C}} p_i^{\mathrm{C}} \tag{9}$$

In the next step, the parameter of interest $\theta$ is introduced by $p_i^{\mathrm{T}} = \theta_i p_i^{\mathrm{C}}$. This leads to

$$lf_{\mathrm{PL}}(x_i|\theta_i, p_i^{\mathrm{C}}) = x_i^{\mathrm{T}} \log(\theta_i) + (x_i^{\mathrm{T}} + x_i^{\mathrm{C}}) \log(p_i^{\mathrm{C}}) - (n_i^{\mathrm{T}} \theta_i + n_i^{\mathrm{C}}) p_i^{\mathrm{C}} \tag{10}$$

Equation (10) is also dependent, besides on $\theta_i$, on the baseline parameter $p_i^{\mathrm{C}}$, the *nuisance* parameter. The nuisance parameter is not in focus here but is important for a valid inference. To follow the PL-approach, $p_i^{\mathrm{C}}$ will be replaced by its *profile maximum likelihood estimator* (PMLE). In this case the PMLE can be derived in a closed form solution, by

$$p_i^{\mathrm{C}} = \frac{x_i^{\mathrm{T}} + x_i^{\mathrm{C}}}{n_i^{\mathrm{T}} \theta_i + n_i^{\mathrm{C}}}$$

In the last step, the PMLE replaces $p_i^{\mathrm{C}}$ in (10), and after simplification, leads to

$$lf_{\mathrm{PL}}(x_i|\theta_i) = x_i^{\mathrm{T}} \log(\theta_i) - (x_i^{\mathrm{T}} + x_i^{\mathrm{C}}) \log(n_i^{\mathrm{T}} \theta_i + n_i^{\mathrm{C}}) \tag{11}$$

Indeed this is an attractive *profile-likelihood* which can be handled easily. From (11) the density function of the PL model can be derived as

$$f_{\mathrm{PL}}(x_i|\theta_i) = \mathrm{e}^{-x_i} \frac{\left(\dfrac{n_i^{\mathrm{T}} \theta_i x_i}{n_i^{\mathrm{C}} + n_i^{\mathrm{T}} \theta_i}\right)^{x_i^{\mathrm{T}}} \left(\dfrac{n_i^{\mathrm{C}} x_i}{n_i^{\mathrm{C}} + n_i^{\mathrm{T}} \theta_i}\right)^{x_i^{\mathrm{C}}}}{x_i^{\mathrm{T}}! x_i^{\mathrm{C}}!}$$

Following the way of the precedent models, the mixture distribution has the form:

$$f(x_i|P) = \sum_{j=1}^{m} f_{\mathrm{PL}}(x_i|\theta_i) q_j$$

## 3. COMPARING PROFILE AND APPROXIMATE LIKELIHOOD

Here, we elaborate on the similarities and differences between the approximate likelihood developed in Section 2.2 and the profile likelihood of Section 2.4. Let us consider the



DOI: 10.1002/sim

*approximate log-likelihood* which is given for the $i$th study—up to an additive constant—as

$$\mathrm{AL}_i(\phi_i) = -\tfrac{1}{2}(\phi_i - \hat{\phi}_i)^2/\hat{\sigma}_i^2 \tag{12}$$

where $\hat{\phi}_i = \log(x_i^{\mathrm{T}}/n_i^{\mathrm{T}}) - \log(x_i^{\mathrm{C}}/n_i^{\mathrm{C}})$ is the estimated log-relative risk in the $i$th study and $\hat{\sigma}_i^2 = 1/x_i^{\mathrm{T}} + 1/x_i^{\mathrm{C}}$ the associated estimated variance. The *profile log-likelihood* in the $i$th study is provided as

$$\mathrm{PL}_i(\phi_i) = x_i^{\mathrm{T}}\phi_i - (x_i^{\mathrm{T}} + x_i^{\mathrm{C}})\log(n_i^{\mathrm{T}}\mathrm{e}^{\phi_i} + n_i^{\mathrm{C}}) \tag{13}$$

where again $\phi_i = \log(\theta)$ is the log-relative risk in the $i$th study.

### 3.1. The likelihoods for centre-specific parameters

Comparing both log-likelihoods for centre-specific parameters $\phi_i$ is in principle identical in doing so for only one centre. Therefore, we can drop the index $i$. We have that

$$\mathrm{PL}(\phi) \approx \mathrm{PL}(\hat{\phi}) + (\phi - \hat{\phi})\mathrm{PL}'(\hat{\phi}) + \tfrac{1}{2}(\phi - \hat{\phi})^2\mathrm{PL}''(\hat{\phi}) \tag{14}$$

$$= \mathrm{PL}(\hat{\phi}) + \tfrac{1}{2}(\phi - \hat{\phi})^2\mathrm{PL}''(\hat{\phi}) \tag{15}$$

using a second-order Taylor expansion around $\hat{\phi}$ and that $\mathrm{PL}'(\hat{\phi}) = 0$. One easily verifies that $\mathrm{PL}''(\hat{\phi}) = -x^{\mathrm{T}}x^{\mathrm{T}}/(x^{\mathrm{T}} + x^{\mathrm{C}}) = -1/\hat{\sigma}_i^2$, showing that

$$\mathrm{PL}(\phi) \approx \mathrm{PL}(\hat{\phi}) + \mathrm{AL}(\phi) \tag{16}$$

so that profile and approximate log-likelihood become identical in a neighbourhood of the maximum likelihood estimator with both log-likelihoods sharing the same curvature. Of course, both log-likelihoods are maximized by the same estimator.

### 3.2. The likelihoods for restricted parameters

Comparison of log-likelihoods starts to become different when parameters are restricted such as in the situation of the hypothesis of *homogeneity*, e.g. $\phi_1 = \phi_2 = \cdots = \phi_k = \phi$. The *approximate log-likelihood* becomes—using independence of the $k$ studies

$$\mathrm{AL}(\phi) = \sum_i \mathrm{AL}_i(\phi) = -\frac{1}{2}\sum_i (\phi - \hat{\phi}_i)^2/\hat{\sigma}_i^2 \tag{17}$$

and the *profile log-likelihood* takes the form

$$\mathrm{PL}(\phi) = \sum_i \mathrm{PL}_i(\phi) = \sum_i x_i^{\mathrm{T}}\phi - (x_i^{\mathrm{T}} + x_i^{\mathrm{C}})\log(n_i^{\mathrm{T}}\mathrm{e}^{\phi} + n_i^{\mathrm{C}}) \tag{18}$$

Figure 2 shows both log-likelihoods for the example. Note that $\mathrm{AL}(\phi)$ is maximized for $\hat{\phi}_w = \sum_i \hat{\phi}_i/\hat{\sigma}_i^2 / \sum_i 1/\hat{\sigma}_i^2$ [5]. In addition, the curvature is given as

$$\mathrm{AL}''(\phi) = -\sum_i 1/\hat{\sigma}_i^2 = -\sum_i \frac{x_i^{\mathrm{T}}x_i^{\mathrm{C}}}{x_i^{\mathrm{T}} + x_i^{\mathrm{C}}} \tag{19}$$





DOI: 10.1002/sim

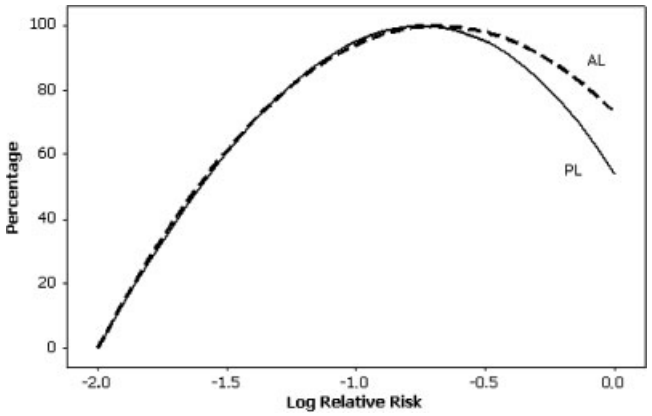


Figure 2. Comparison of the profile and approximate log-likelihood from the meta-analysis.

The profile log-likelihood is maximized for $\hat{\theta} = \exp(\hat{\phi})$ satisfying

$$\hat{\theta} = \frac{\sum_i x_i^{\mathrm{T}} n_i^{\mathrm{C}} / (n_i^{\mathrm{T}} \hat{\theta} + n_i^{\mathrm{C}})}{\sum_i x_i^{\mathrm{C}} n_i^{\mathrm{T}} / (n_i^{\mathrm{T}} \hat{\theta} + n_i^{\mathrm{C}})} \tag{20}$$

and has curvature

$$\mathrm{PL}''(\phi) = \sum_i -(x_i^{\mathrm{T}} + x_i^{\mathrm{C}}) \frac{n_i^{\mathrm{T}} n_i^{\mathrm{C}} \mathrm{e}^{\phi}}{(n_i^{\mathrm{T}} \mathrm{e}^{\phi} + n_i^{\mathrm{C}})^2} \tag{21}$$

Approximate and profile log-likelihood are not only maximized at different parameter values, the curvature of the profile log-likelihood at the maximum likelihood estimate is different from the curvature of the approximate log-likelihood. To explore this point in more detail, let us assume that the trial is balanced e.g. $n_i^{\mathrm{T}} = n_i^{\mathrm{C}}$ for all $i$. Then, the PMLE is available in closed form $\hat{\theta} = \sum_i x_i^{\mathrm{T}} / \sum_i x_i^{\mathrm{C}}$ and the curvature at $\hat{\phi}$ is simply

$$\begin{aligned} \mathrm{PL}''(\hat{\phi}) &= \sum_i -(x_i^{\mathrm{T}} + x_i^{\mathrm{C}}) \frac{\dfrac{\sum_i x_i^{\mathrm{T}}}{\sum_i x_i^{\mathrm{C}}}}{\left( \dfrac{\sum_i x_i^{\mathrm{T}}}{\sum_i x_i^{\mathrm{C}}} + 1 \right)^2} \\ &= -\frac{\sum_i x_i^{\mathrm{T}} \sum_i x_i^{\mathrm{C}}}{\sum_i (x_i^{\mathrm{T}} + x_i^{\mathrm{C}})} \end{aligned} \tag{22}$$

It is remarkable that a general comparison between these two curvatures is possible, indicating a more precise estimator based upon the profile likelihood.





DOI: 10.1002/sim

*Theorem*
Let the all centres involved in the MAIPD be balanced, e.g. $n_i^{\mathrm{T}} = n_i^{\mathrm{C}}$ for all $i$. Then,

$$\mathrm{PL}''(\hat{\phi}) \leqslant \mathrm{AL}''(\phi)$$

and

$$\widehat{\mathrm{var}}(\hat{\phi}) \leqslant \widehat{\mathrm{var}}(\hat{\phi}_w)$$

*Proof*
We show

$$\frac{\sum_i x_i^{\mathrm{T}} \sum_i x_i^{\mathrm{C}}}{\sum_i (x_i^{\mathrm{T}} + x_i^{\mathrm{C}})} \geqslant \sum_i \frac{x_i^{\mathrm{T}} x_i^{\mathrm{C}}}{x_i^{\mathrm{T}} + x_i^{\mathrm{C}}}$$

or equivalently

$$\frac{\frac{1}{k}\sum_i x_i^{\mathrm{T}} \frac{1}{k}\sum_i x_i^{\mathrm{C}}}{\frac{1}{k}\sum_i (x_i^{\mathrm{T}} + x_i^{\mathrm{C}})} \geqslant \frac{1}{k}\sum_i \frac{x_i^{\mathrm{T}} x_i^{\mathrm{C}}}{x_i^{\mathrm{T}} + x_i^{\mathrm{C}}}$$

This follows from the fact that the function $g(y, z) = yz/(y + z)$, defined for $y>0$, $z>0$ is concave which is proved by showing that the *Hessian* of $g(y, z)$

$$\begin{pmatrix} -2z & y - z \\ z - y & -2y \end{pmatrix} \Big/ (y + z)^3$$

is negative definite. This ends the proof. □

## 4. ESTIMATING THE MIXING DISTRIBUTION WITH THE EM ALGORITHM

Parameters of the mixing distribution are estimated by means of the maximum likelihood estimation (MLE) via the EM algorithm. Suppose there are $m$ components in the population with relative risk $\theta_1, \theta_2, \ldots, \theta_m$. We introduce a latent indicator variable $z_{ij}$, which is 1 if trial $i$ belongs to component $j$, and 0 otherwise. Suppose further that the outcome is

$$x_i = \begin{cases} \hat{\phi}_i = \log\left(\dfrac{x_i^{\mathrm{T}}}{n_i^{\mathrm{T}}}\right) - \log\left(\dfrac{x_i^{\mathrm{C}}}{n_i^{\mathrm{C}}}\right) & \text{in the AL model} \\ (x_i^{\mathrm{T}}, n_i^{\mathrm{T}}, x_i^{\mathrm{C}}, n_i^{\mathrm{C}})' & \text{otherwise} \end{cases}$$

and, conditional on membership in component $j$, has density $f(x_i|\theta_j)$, then the unconditional, joint density of $(x_i, z_{ij})$ is given as

$$\prod_{j=1}^{m} [q_j f(x_i|\theta_j)]^{z_{ij}} \tag{23}$$





DOI: 10.1002/sim

where the product in (23) is taken over all components. The full-sample log likelihood becomes

$$\sum_{i=1}^{k}\sum_{j=1}^{m} z_{ij}\log[q_j f(x_i|\theta_j)] = \sum_{i=1}^{k}\sum_{j=1}^{m} z_{ij}\log[q_j] + \sum_{i=1}^{k}\sum_{j=1}^{m} z_{ij}\log[f(x_i|\theta_j)] \tag{24}$$

An advantage is that in (24) the $q$'s and $\theta$'s can be maximized separately. In the EM algorithm, the unobserved indicator $z_{ij}$ is replaced by its expected value, conditional on current values of $\theta_j$ and $q_j$ $(j = 1, \ldots, m)$ leading to the $E$-step

$$e_{ij} = E(z_{ij}|x_i, q_j, \theta_j) = \frac{q_j f(x_i|\theta_j)}{\sum_{l=1}^{m} q_l f(x_i|\theta_l)} \tag{25}$$

Replacing $e_{ij}$ for $z_{ij}$ in (24) gives the *expected log-likelihood*

$$\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log[q_j f(x_i|\theta_j)] = \sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log q_j + \sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log f(x_i|\theta_j)$$

to be maximized in $q_j$ and $\theta_j$ (M-step). The maximization of $\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log q_j$ gives the conventional result for each model, as

$$q_j^{\text{new}} = \frac{\sum_{i=1}^{k} e_{ij}}{k}$$

The maximization of $\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log[f(x_i, \theta_j)]$, leading to the M-step, is dependent on the density of the model under consideration. Details for the M-steps in these three models are provided in Appendix A.

For the PL-approach, there is no closed solution form available in the M-step. But the provided fix-point iteration (A8) converges generally very fast to the non-parametric profile maximum likelihood estimator (NPMLE). One *important* point here is that in the case of homogeneity $m = 1$, if in the first iteration $\theta = 1$ is used as initial value, then the one-step-estimator will coincide with the Mantel–Haenszel estimator

$$\theta_{\text{MH}} = \frac{\sum_{i=1}^{k} x_i^{\text{T}} n_i^{\text{C}} (n_i^{\text{T}} + n_i^{\text{C}})^{-1}}{\sum_{i=1}^{k} x_i^{\text{C}} n_i^{\text{T}} (n_i^{\text{T}} + n_i^{\text{C}})^{-1}}$$

so that the profile likelihood estimator might be viewed as an *extension* of the Mantel–Haenszel approach.

After the estimation has been carried out, it is possible to classify the studies into the found components, because we can interpret the $e_{ij}$ (25) as *a posteriori* probability of study $i$ belonging to component $j$. Hence, one could classify study $i$ into that component $j$ for which (25) is the largest among all components.

### *4.1. Choice of the mixture model*

The number of components is generally unknown and has to be estimated. Several criteria are available. In this contribution we consider two of these. The first is the Bayesian information criterion (BIC), which goes back to Schwarz [17] and has been further discussed by many authors





DOI: 10.1002/sim

including McLachlan and Peel [18], which penalized the log-likelihood $l(P)$ by $d\log(k)$ and is given as

$$\mathrm{BIC} = 2l(P) - d\log(k)$$

with $d =$ number of estimated parameters

$$d = (2m - 1) \quad \text{for the AL and PL model}$$

$$d = (3m - 1) \quad \text{for the ML model}$$

According to this criterion the number of components $m$ is chosen with the largest BIC-value. This criterion is frequently recommended as a guideline for selecting the number of components [18]. The second criterion is the non-parametric maximum likelihood criterion (NPLME) descending from the gradient function (see References [19, 20]), which is defined as

$$d(\theta, P) = \frac{1}{k}\sum_{i=1}^{k}\frac{f(x_i|\theta)}{f(x_i|P)} \tag{26}$$

and derived from the directional derivative. An important theorem in the theory of non-parametric mixture models is the *general mixture maximum likelihood theorem* stating that, if the gradient function is bounded above by one for all $\theta$ in the parameter space, then $\hat{P}$ is the unique maximum of the likelihood function. This theorem is very useful in verifying an estimated mixing distribution for optimality.

## 5. ANALYSIS FOR THE MAIPD ON SELECTIVE TRACT DECONTAMINATION

In this section our objective is to analyse the relative risk structure of the MAIPD provided in Table I. The results from the AL and PL models are given in Table II. Both methods find two mixture components as the largest number of components. The PL classified twelve studies and the AL 13 studies to the first component with an estimated relative risk of 0.56. Consequently, in these studies the risk of respiratory infection is almost halved in comparison to the control group. The second component estimated a relative risk of 0.23 (PL) and 0.25 (AL). Apparently, the estimators from both models are very close together. One difference lies in the BIC. In the fixed effect model the BIC estimated only one component, whereas in the PL model two components were chosen. In the ML many more components were found Table III. In this approach a maximum of six components were observed. The 4th and 5th component estimate nearly the same relative risk, only the baseline is different. With the BIC, four components were selected as appropriate number of components. One important difference between the ML model and the PL and AL model is found in the way the classification of studies into the associated components is done, see Figures 3 and 4. The study allocation of the AL model is similar to the PL model, only the first study is allocated differently. In Figure 3, it can be seen that, for example, the 3rd and the 16th study are very close to the first component line. This means that these studies have the same or similar relative risk as the first component in the ML model, although the studies are classified into the second component. The reason for this misallocation lies in the influence of the baseline heterogeneity on the estimation of effect heterogeneity. In contrast, Figure 4 shows that all studies





DOI: 10.1002/sim

Table II. Results of MAIPD (Table I) from the approximate and profile likelihood model ($H$ is number of studies which belong to the respective component).

| Profile-likelihood model | | | Approximate likelihood model | | |
|---|---|---|---|---|---|
| Comp. | 1. | | Comp. | 1. | |
| $\theta$ | 0.473867 | | $\theta$ | 0.501208 | |
| $q$ | 1.000000 | | $q$ | 1.000000 | |
| $H$ | 22 | | $H$ | 22 | |
| Log-$L$. $= -116.606398$ | | | Log-$L$. $= -20.638539$ | | |
| max GF $= 2.156801$ | | | max GF $= 1.388575$ | | |
| BIC $= -236.3038$ | | | BIC $= -44.3681$ | | |
| Comp. | 1. | 2. | Comp. | 1. | 2. |
| $\theta$ | 0.562660 | 0.231091 | $\theta$ | 0.559415 | 0.258274 |
| $q$ | 0.615078 | 0.384922 | $q$ | 0.678406 | 0.321594 |
| $H$ | 12 | 10 | $H$ | 13 | 9 |
| Log-$L$. $= -113.359637$ | | | Log-$L$. $= -19.418858$ | | |
| max GF $= 1.000000$ | | | max GF $= 1.000000$ | | |
| BIC $= -235.9924$ | | | BIC $= -48.1108$ | | |

are allocated on the basis of the treatment effect in the PL and AL model (which is the major interest of the practitioner).

## 6. SIMULATION STUDY

In this section, all three models are compared by means of simulation studies.

### *6.1. Two component effect heterogeneity*

It is assumed that in the first simulation experiment the population of interest consists of two clusters. The clusters are represented by the mixing distribution $P = \binom{0.5\quad 1.5}{0.5\quad 0.5}$. Both components receive an identical weight of 0.5. The first component has a relative risk of 0.5 and the second of 1.5. To mimic baseline variation, the baseline risks $p_1^{\mathrm{C}}, \ldots, p_k^{\mathrm{C}}$ were generated from a uniform distribution from 0.1 to 0.66. The parameter $p_i^{\mathrm{T}}$ depends on the component the $i$th study belongs to. If the $i$th study belongs to the first component, then $p_i^{\mathrm{T}} = \theta_1 p_i^{\mathrm{C}} = 0.5 p_i^{\mathrm{C}}$, otherwise $p_i^{\mathrm{T}} = \theta_2 p_i^{\mathrm{C}} = 1.5 p_i^{\mathrm{C}}$. In this case the weights are equal, so that component membership of each study is generated by means of a Bernoulli distribution with 0.5 event probability. The sample size $n_i^{\mathrm{T}}$ and $n_i^{\mathrm{C}}$ were generated from a Poisson distribution with parameter 100. Poisson variates $X_i^{\mathrm{T}}$ with parameters $n_i^{\mathrm{T}}$ and $p_i^{\mathrm{T}}$ and Poisson variates $X_i^{\mathrm{C}}$ with parameters $n_i^{\mathrm{C}}$ and $p_i^{\mathrm{C}}$ were drawn for each study $i$, $i = 1, \ldots, k$. In this case the number of studies was chosen to be $k = 100$. For reasons of comparability only a two component mixture was estimated for all three models. The procedure was replicated 1000 times. From this replication the mean and variance of each component were computed. The results of this constellation are provided in Figure 5. The first component of the ML model is considerably over-estimated. Note that actually the true relative risk is not captured





DOI: 10.1002/sim

Table III. Results of MAIPD (Table I) using the multi-level model ($H$ is number of studies which belong to the respective component).

| Comp. | 1. | | | | | |
|---|---|---|---|---|---|---|
| $\alpha$ | −1.228386 | | | | | |
| $\theta$ | 0.467574 | | | | | |
| $q$ | 1.000000 | | | | | |
| $H$ | 22 | | | | | |
| Log-$L$. = −243.585572 | | | | | | |
| max GF = $2.78 \times 10^{12}$ | | | | | | |
| BIC = −490.2622 | | | | | | |

| Comp. | 1. | 2. |
|---|---|---|
| $\alpha$ | −1.660445 | −0.659691 |
| $\theta$ | 0.548765 | 0.418367 |
| $q$ | 0.501842 | 0.498158 |
| $H$ | 11 | 11 |
| Log-$L$. = −171.325585 | | |
| max GF = 6344.888224 | | |
| BIC = −351.9243 | | |

| Comp. | 1. | 2. | 3. | 4. |
|---|---|---|---|---|
| $\alpha$ | −1.576561 | −0.778285 | −0.044227 | −1.944845 |
| $\theta$ | 0.635286 | 0.379265 | 0.508046 | 0.229756 |
| $q$ | 0.273606 | 0.413516 | 0.091149 | 0.221729 |
| $H$ | 6 | 9 | 2 | 5 |
| Log-$L$. = −140.834193 | | | | |
| max GF = 2.206511 | | | | |
| BIC = −303.3057 | | | | |

| Comp. | 1. | 2. | 3. | 4. | 5. | 6. |
|---|---|---|---|---|---|---|
| $\alpha$ | −1.592677 | −0.835141 | −0.044119 | −1.957214 | −0.685773 | −1.528249 |
| $\theta$ | 0.697564 | 0.475019 | 0.508104 | 0.228241 | 0.212911 | 0.343012 |
| $q$ | 0.224330 | 0.201705 | 0.091086 | 0.211572 | 0.213049 | 0.058259 |
| $H$ | 5 | 3 | 2 | 5 | 6 | 1 |
| Log-$L$. = −138.957720 | | | | | | |
| max GF = 1.000000 | | | | | | |
| BIC = − 311.9169 | | | | | | |

by the confidence interval. In contrast the second component is under-estimated. In the other two models the true distribution is recovered.

### 6.2. *Under effect homogeneity*

The next simulation study investigated the situation of effect homogeneity. In this case we used a bootstrap simulation (see Reference [21]). The differences are expected in the sparsity case, where the number of observations and probands are rare. For this we used a special sparsity meta-analysis, namely CALGB study adopted from Lipsitz *et al.* [22]. The main settings for the simulation: $p_i^{\mathrm{C}}$, $n_i^{\mathrm{T}}$ and $n_i^{\mathrm{C}}$ stem from the sparsity study; $\theta$ is predetermined and fixed for all studies $i = 1, \ldots, k$; $X_i^{\mathrm{T}}$ with parameters $n_i^{\mathrm{T}} \times \theta p_i^{\mathrm{C}}$ and $X_i^{\mathrm{C}}$ with parameters $n_i^{\mathrm{C}} \times p_i^{\mathrm{C}}$ were generated from an Poisson distribution. Figure 6 shows the bias of 30 values for $\theta$ in the interval from 0.1 to 0.99 for





DOI: 10.1002/sim

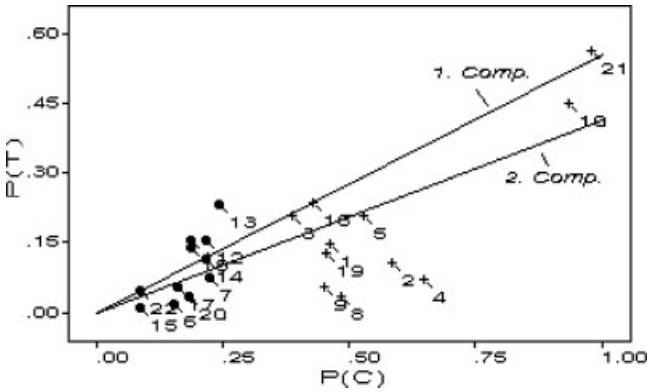


Figure 3. Study allocation to the components for the multilevel model (circle ≙ 1. comp., cross ≙ 2. comp.).

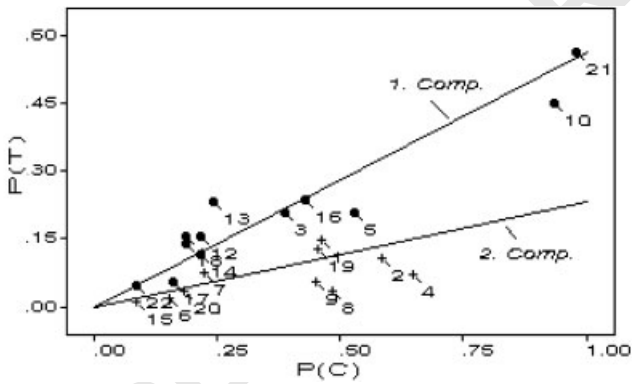


Figure 4. Study allocation to components for the profile-likelihood model (circle ≙ 1. comp., cross ≙ 2. comp.).

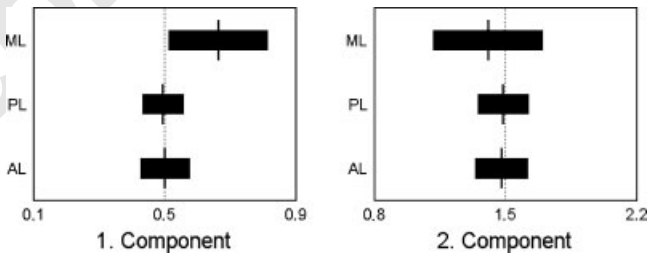


Figure 5. Results of a simulation study of a two component mixture (with baseline heterogeneity) for the three models AL, PL, ML to estimate the predetermined mixing components {0.5, 1.5} with weights {0.5, 0.5}. displayed are the means with 95 per cent confidence intervals for each estimated component.



DOI: 10.1002/sim

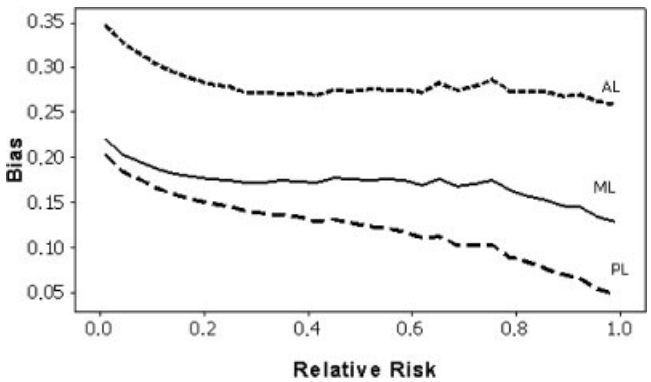


Figure 6. Bias of ML, Pl, AL-model in simulation study under effect homogeneity.

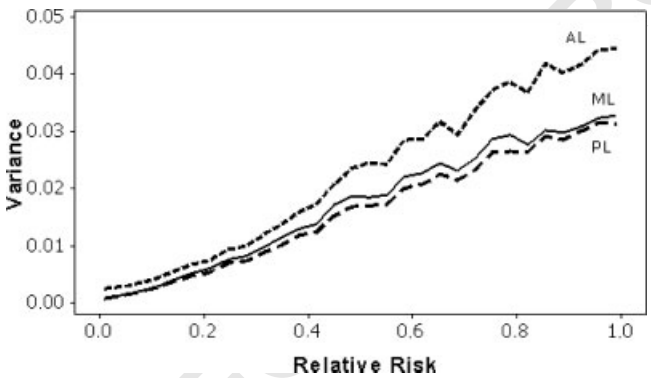


Figure 7. Variance of ML, Pl, AL-model estimator in simulation study under effect homogeneity.

all the three models. In this Figure the PL-model estimator has the smallest bias. In contrast, the other two models have a considerably over-estimated the true value. One reason could be that the estimator of the ML model, here the crude risk ratio estimator

$$\hat{\theta}_{\text{crude}} = \frac{\sum_{i=1}^{k} x_i^{\text{T}} \sum_{i=1}^{k} n_i^{\text{C}}}{\sum_{i=1}^{k} n_i^{\text{T}} \sum_{i=1}^{k} x_i^{\text{C}}}$$

adjusts for a potential centre effect. Also, the weighted estimator used in the AL model, where the weight originates from the inverse of the variance of log relative risk, might not be appropriate in this situation.

Drawing the attention to the variance (Figure 7), the AL estimator has a slightly larger value than the ML and PL estimators (which is consistent with the theorem of Section 3.2), whereas the values of the variance of ML and PL model are very close.



DOI: 10.1002/sim

## 7. DISCUSSION

All three models discussed here are designed to estimate the treatment effect of a meta-analysis with individually pooled data. Although this is their common aspect, there are also important differences between them. One important difference lies in the way the nuisance or baseline parameter is treated. In the AL model the baseline is integrated into the individual log relative risk. The specific aspect of the PL approach is that the nuisance parameter is integrated into the likelihood in such a way that the occurring likelihood, the profile likelihood, depends only on the parameter of interest. Here, this method provides a simple form of the profile log-likelihood function. In contrast, the ML method does not eliminate the nuisance parameter, but estimates it as a separate parameter. The results in Table III and the simulation study in Figure 5 show that this model looses power when estimating baseline heterogeneity. Furthermore, the allocation of studies or centres to the mixed components is also dependent on the baseline parameter. In other words, the baseline parameter has a very strong influence on estimating the treatment effect. In the situation of an increased baseline heterogeneity it can happen, like in the simulation study in Figure 5, that the result of estimating the treatment effect heterogeneity is confounded by the existing strong baseline heterogeneity. Consequently, a substantial disadvantage of the ML model can be seen in the handling of the baseline parameter. The PL-model has the advantage of integrating the estimation of the baseline parameter into the estimation of treatment effect.

It should be mentioned that the profile method is a conventional way to deal with nuisance parameters, but by no means the only way. In the ideal case (see Reference [23]), parameter of interest and nuisance parameter are orthogonal, that is, the joint likelihood $\pounds(\theta, p^{\mathrm{C}}) = \pounds_1(\theta)\pounds_2(p^{\mathrm{C}})$ factors into likelihood depending only on $\theta$ and $p^{\mathrm{C}}$, respectively. For the ease of discussion only one trial is considered, though generalizations are straight-forward. Write the joint likelihood $\exp(-p^{\mathrm{T}}n^{\mathrm{T}})(p^{\mathrm{T}}n^{\mathrm{T}})^{x^{\mathrm{T}}} \times \exp(-p^{\mathrm{C}}n^{\mathrm{C}})(p^{\mathrm{C}}n^{\mathrm{C}})^{x^{\mathrm{C}}}$ as product of $\pounds_1(\theta) = (n^{\mathrm{T}}\theta/(n^{\mathrm{C}} + n^{\mathrm{T}}\theta))^{x^{\mathrm{T}}}$ $(n^{\mathrm{C}}/(n^{\mathrm{C}} + n^{\mathrm{T}}\theta))^{x^{\mathrm{C}}}$ and $\pounds_2(\eta_n) = \exp(-\eta_n)\eta_n^{x^{\mathrm{T}}+x^{\mathrm{C}}}$, where $\theta$ is the risk ratio and $\eta_n = n^{\mathrm{T}}p^{\mathrm{T}} + n^{\mathrm{C}}p^{\mathrm{C}}$. In case that the trial is balanced $\eta_n = n^{\mathrm{T}}p^{\mathrm{T}} + n^{\mathrm{C}}p^{\mathrm{C}} = \eta(n^{\mathrm{T}} + n^{\mathrm{C}})$, and $\theta$ and $\eta = p^{\mathrm{T}} + p^{\mathrm{C}}$ are *orthogonal*. In the case of orthogonality, one can solely base inference on $\pounds_1(\theta)$, and the profile likelihood is identical to $\pounds_1(\theta)$ which is also a true likelihood. If the trial is unbalanced, the transformation $\eta_n = n^{\mathrm{T}}p^{\mathrm{T}} + n^{\mathrm{C}}p^{\mathrm{C}}$ necessarily incorporates the known, trial-specific sample size parameters, but $\pounds_1(\theta)$ will remain identical. Alternatively, one may base inference on the likelihood conditional on the sufficient statistic $x = x^{\mathrm{T}} + x^{\mathrm{C}}$ for the nuisance parameter, and, although this is by no means in generality the case, it does turn out again to be $\pounds_1(\eta)$ (see for a more general discussion Reference [23] or [24]). Yet, another way in dealing with the nuisance parameter is suggested in Reference [25]. It is suggested to use as effect measure the odds ratio and base inference on the full, product-binomial likelihood—without eliminating the nuisance parameter. In a second step, inference is based upon the distribution of $X^{\mathrm{C}}$—conditional upon $X^{\mathrm{T}}$ and $X^{\mathrm{C}}$. The occurring non-central hypergeometric distribution is a function of the odds ratio only, so that the associated likelihood is free of the nuisance parameter. This appears to be an attractive approach and—despite the complex character of the non-central hypergeometric likelihood—should be analysed in further depth and compared with the profile approach in future work.

The second issue is the modelling of the occurrence of heterogeneity. Very often, especially in random effects models, it is assumed that the parameter of interest has a continuous distribution (see Reference [12] or [5]). We assume a discrete distribution for the mixing distribution (see in detail Reference [19]) which explains the unobserved heterogeneity and occurs in a natural way





DOI: 10.1002/sim

as the marginal distribution of the unobserved covariate. Van Houwelingen *et al.* [25] follow this route as well by using a discrete mixture of non-central hypergeometric distributions. However, Van Houwelingen *et al.* [25] find it more desirable to have a smooth mixing distribution and use a normal distribution in estimating the true mixing distribution. In our view, the problem lies in the fact that true model for the mixing distribution is unknown, and if it is left unspecified, the resulting estimate is necessarily discrete. If, in fact, the true mixing distribution is continuous, we would argue that the discretely estimated mixing distribution would provide a reasonable approximation, at least it has a likelihood always as large as the corresponding one of the continuous mixing distribution. Note also that modelling a discrete mixture distribution includes the potential case of effect homogeneity. Frequently in MAIPD, only one component is found, so that effect heterogeneity can be excluded without further statistical testing.

In some MAIPD study-covariates such as treatment modifications, time of study, etc. are available. Some heterogeneity might be explained by means of these covariates and it is possible to extend the modelling to incorporate the covariates into each of the three models (see also Reference [26] or, in particular for the PL-method, Reference [27]).

The main result of this paper can be seen in the fact that the PL model can be considered as a clear alternative compared to the AL method. When comparing the PL method with the ML model, the role of the baseline heterogeneity must be considered. When a strong baseline heterogeneity occurs, as can be seen in the example in Figure 5, the PL method is preferred over the ML model. The ML model is more adequate for estimating the treatment effect if baseline homogeneity exists. However, in meta-analysis, baseline homogeneity rarely occurs and is therefore a strong assumption.

One limiting assumption of the PL model is the Poisson distribution for the study specific event counts. This might be realistic if the MAIPD consists of event counts within person-times, although study data of the MAIPD are often given as event counts per number under risk, so that a binomial distribution might be more appropriate.

The Poisson distribution is, however, a reasonable approximation of the binomial distribution and an excellent approximation if the event counts are sparse. In principle, it is possible to apply the PL framework to the binomial likelihood, but in this case the PL function becomes complex and is difficult to evaluate. Therefore, it looses its attractivity to practitioners. The results obtained here are derived from a PhD thesis and detailed comparisons can be found there (see Reference [4]). In particular, it could be demonstrated that the effect of the misspecification of the likelihood is minor and in most cases could be ignored.

## APPENDIX A: DETAILS ON THE M-STEP IN THE EM ALGORITHM FOR THE THREE MODELS

Here, the various forms of the M-step are derived for the three models. All expected log-likelihoods (ignoring only data dependent terms) and the associated estimators are given as follows:

- Approximate likelihood model

$$\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}\log[f(x_i|\theta_j)] = \sum_{i=1}^{k}\sum_{j=1}^{m} -\frac{e_{ij}(x_i-\log(\theta_i))^2}{2\hat{\sigma}_i^2} \tag{A1}$$



DOI: 10.1002/sim

Estimator:

$$\hat{\theta}_j = \mathrm{e}^{\sum_{i=1}^{k} e_{ij} x_i / \hat{\sigma}_i^2 / \sum_{i=1}^{k} e_{ij} / \hat{\sigma}_i^2} \tag{A2}$$

- Multi-level-model

$$\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij} \log[f(x_i|\alpha_j, \beta_j)] = \sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij}(x_i^{\mathrm{C}}\alpha_j - n_i^{\mathrm{C}}\mathrm{e}^{\alpha_i} + x_i^{\mathrm{T}}(\alpha_j + \beta_j) - n_i^{\mathrm{T}}\mathrm{e}^{\alpha_i+\beta_i}) \tag{A3}$$

Estimators:

$$\hat{\alpha}_j = \log\left(\frac{\sum_{i=1}^{k} e_{ij} x_i^{\mathrm{C}}}{\sum_{i=1}^{k} e_{ij} n_i^{\mathrm{C}}}\right) \tag{A4}$$

$$\hat{\beta}_j = \log\left(\frac{\sum_{i=1}^{k} e_{ij} x_i^{\mathrm{T}} \sum_{i=1}^{k} e_{ij} n_i^{\mathrm{C}}}{\sum_{i=1}^{k} e_{ij} n_i^{\mathrm{T}} \sum_{i=1}^{k} e_{ij} x_i^{\mathrm{C}}}\right) \tag{A5}$$

$$\hat{\theta}_j = \frac{\sum_{i=1}^{k} e_{ij} x_i^{\mathrm{T}} \sum_{i=1}^{k} e_{ij} n_i^{\mathrm{C}}}{\sum_{i=1}^{k} e_{ij} n_i^{\mathrm{T}} \sum_{i=1}^{k} e_{ij} x_i^{\mathrm{C}}} \tag{A6}$$

The effect parameter $\hat{\beta}_j$ is in this case the log relative risk.

- Profile-likelihood model

$$\sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij} \log[f(x_i|\theta_j)] = \sum_{i=1}^{k}\sum_{j=1}^{m} e_{ij} x_i^{\mathrm{T}} \log(\theta_i) - e_{ij} \log(n_i^{\mathrm{T}}\theta_i + n_i^{\mathrm{C}}) \tag{A7}$$

Estimator:

$$\hat{\theta}_j = \frac{\sum_{i=1}^{k} e_{ij} x_i^{\mathrm{T}} n_i^{\mathrm{C}} (n_i^{\mathrm{T}}\hat{\theta}_j + n_i^{\mathrm{C}})^{-1}}{\sum_{i=1}^{k} e_{ij} x_i^{\mathrm{C}} n_i^{\mathrm{T}} (n_i^{\mathrm{T}}\hat{\theta}_j + n_i^{\mathrm{C}})^{-1}} := \Gamma(\theta_j) \tag{A8}$$

ACKNOWLEDGEMENTS

We are very grateful to the Editor, the Associate Editor, and two referees for their helpful comments.

REFERENCES

1. Olkin I. Meta-analysis: reconciling the results of independent studies. *Statistics in Medicine* 1995; **14**:457–472.
2. Jones DR. Meta-analysis: weighing the evidence. *Statistics in Medicine* 1995; **14**:137–149.
3. Fleiss JL. The statistical basis of meta-analysis. *Statistical Methods in Medical Research* 1993; **2**:121–145.
4. Kuhnert R. Untersuchung von verschiedenen Modellierungen der Heterogenität in multizentrischen Studien. *Dissertation*, Charité Medical School to be published under http://www.diss.fu-berlin.de/2005/202 [31 August 2005].



DOI: 10.1002/sim

5. Van Houwelingen HC, Arends LR, Stijnen T. Tutorial in Biostatistics. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
6. Aitkin M. Meta-analysis by random effect modelling in generalized linear models. *Statistics in Medicine* 1999; **18**:2343–2351.
7. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. *British Medical Journal* 1993; **307**:525–532.
8. Woodward M. *Epidemiology. Study Design and Data Analysis.* Chapman & Hall/CRC: London, Boca Raton, FL, 1999.
9. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677.
10. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**:173–192.
11. Whitehead A. *Meta-Analysis of Controlled Clinical Trials.* Wiley: Chichester, 2002.
12. DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
13. Laird NM. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* 1978; **73**:805–811.
14. Aitkin M. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 1999; **55**:117–128.
15. Aitkin M. Profile likelihood. *Encyclopedia of Biostatistics*. Wiley: Chichester, New York, 1998; 3534–3536.
16. Edwards AWF. *Likelihood*. Johns Hopkin University Press: Baltimore, MD, 1992.
17. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**:461–464.
18. McLachlan G, Peel D. *Finite Mixture Models*. Wiley: New York, 2000.
19. Böhning D. *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis*, *Disease Mapping and Others.* Chapman & Hall/CRC: London, Boca Raton, FL, 2000.
20. Böhning D. The EM algorithm with gradient function update for discrete mixtures with know (fixed) number of components. *Statistics and Computing* 2003; **13**:257–265.
21. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall: London, 1993.
22. Lipsitz SR, Dear KB, Laird NM, Molenberghs G. Tests of homogeneity of risk difference when data are sparse. *Biometrics* 1998; **54**:148–160.
23. Pawitan Y. *In All Likelihood. Statistical Modelling and Inference Using Likelihood.* Clarendon Press: Oxford, 2001.
24. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall: London, 1989.
25. Van Houwelingen HC, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993; **12**:2272–2284.
26. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
27. Rattanasiri S. Modelling covariate information in multicentre studies with binary outcome using profile likelihood. *Dissertation*, Charité Medical School to be published under http://www.diss.fu-berlin.de/2006/344 [15 July 2006].





DOI: 10.1002/sim