Research Report - Forschungsbericht
Final Report - Abschlussbericht

A Biostatistical Investigation and Modelling of Heterogeneous Populations - The Two Sample Case: A Profile Likelihood Mixture Approach - DFG BO 865/6-3,6-4,6-5,6-6 Reporting the Years 2004 and 2005

Dankmar Böhning
Section of Applied Statistics
School of Biological Sciences
Harry Pitt Building, University of Reading
Whiteknights, Reading
E-mail:d.a.w.bohning@reading.ac.uk

July 3, 2006

Contents

1 Summary

A unifying approach to the modelling of the risk ratio for multicenter studies with binary outcome is provided. In these settings, each center is characterized by a baseline or nuisance parameter and a effect or interest parameter. One way of dealing with nuisance parameters is the profile likelihood method for which the basic model is introduced. The profile likelihood method under homogeneity is investigated and the connection to the Mantel-Haenszel approach illuminated. The model is extended to cope with heterogeneity. Unobserved heterogeneity is captured by means of a nonparametric mixture leading to the nonparametric mixture profile likelihood. The gradient function is introduced and the nonparametric profile maximum likelihood estimator (PNMLE) is characterized. The latter can be computed by means of the EM algorithm with gradient function update (EMGFU). Furthermore, modelling of covariate information is introduced. Elements of log-linear modelling are used and ways for finding the profile maximum likelihood estimator including their standard errors are provided. Finally, simulation studies are done to compare the proposed methodology with existing methods and results show clear advantages of the proposed methodology.

Key words: MIX-PRO

Areas of Research: Statistics, particularly Biostatistics and Biometry, Epidemiology

2 Principal Investigators and Partners

- Principal Investigator (responsible for German side)
  - Böhning, Dankmar
  - Professor Dr. rer. nat.
  - Joint Center for Health Sciences and Humanitites Free University Berlin / Humboldt University at Berlin Institute for International Health (until October 2005)
    Chair for Applied Statistics in the Life Sciences, School of Biological Sciences, University of Reading (October 2005 - present)
- Thai Principal Investigator (and full responsible for Thai side)
  - Chukiat Viwatwongkasem
  - Associate Professor, B.Sc. (Math), M.Sc. (Biostatistics)
  - Mahidol University, Faculty of Public Health, Department of Biostatistics, Rajavithi Road 420/1 Bangkok 10400, Thailand
- German Co-Investigators

1. Ekkehart Dietz
   Dr. rer. nat.
   Joint Center for Health Sciences and Humanitites Free University Berlin / Humboldt University at Berlin
   Institute for International Health
2. Ronny Kuhnert
   Dr rer. medic.
   Joint Center for Health Sciences and Humanitites Free University Berlin / Humboldt University at Berlin
   Institute for International Health

- Thai Co-Investigators

    1. Piangchan Rojanavipart
       Associate Professor, M.H.S. (Biostatistics, USA), B.S. (Marine Sc.), Medical Statistics (England), M.Sc. (Biostatistics)
       Mahidol University, Faculty of Public Health, Department of Biostatistics, Rajavithi Road 420/1 Bangkok 10400, Thailand

    2. Dechavudh Nityasuddhi
       Assistant Professor, B.Sc. (Math), M.Sc. (Biostatistics), Medical Statistics (England)
       Mahidol University, Faculty of Public Health, Department of Biostatistics, Rajavithi Road 420/1 Bangkok 10400, Thailand

    3. Sasivimol Rattansiri
       Assistant Professor, B.Sc. (Bio), M.Sc. (Biostatistics),
       Mahidol University, Faculty of Medicine -Ramathibody Hospital, Epidemiological Research Unit, Bangkok 10400, Thailand

3 Achievements during and as a Result of the Project

3.1 Publications

1. Ronny Kuhnert and Dankmar Böhning: A Comparison of Three Different Models For Estimating Relative Risk in Meta-Analysis of Clinical Trials Under Unobserved Heterogeneity. *Statistics in Medicine* (under revision).

2. Dankmar Böhning, Ronny Kuhnert, Chukiat Viwatwongkasem, and Sasivimol Rattanasiri: Nonparametric Profile Likelihood Estimation in Meta-Analysis with Individually Pooled Data. *Statistical Modelling* (submitted).

3.2 PhD-Theses Completed

1. Ronny Kuhnert: Untersuchung von verschiedenen Modellierungen der Heterogenität in multizentrischen Studien. Dissertation at Charité Medical School to be published under http://www.diss.fu-berlin.de-/2005/202 [31 August 2005].

2. Sasivimol Rattanasiri: Modelling Covariate Inofrmation in Multicentre Studies with Binary Outcome using the Profile Likelihood. Dissertation at Charité Medical School to be published under http://www.diss.fu-berlin.de/2006/xx [31 July 2006]

4 Introduction

4.1 The Occurrence of Multicenter Studies

The present contribution aims to provide a *unifying* approach to modelling effect in multicenter studies with binary outcome. Multicenter studies occur in numerous ways and are executed quite frequently. It is said that "there are currently thousands of active multicenter trials designed to evaluate treatment or prevention strategies" (Bryant *et al.* 1998). Among the reasons for mounting a multicenter study are the *need to recruit patients at a faster rate*, the *need to find patients with a rare disease or condition*, or the *desire to increase the generalizability of effect*, since multicenter studies will more likely include heterogenous populations. Pocock (1997) points out that the collaboration of clinical scientists in a multicenter study should lead to increased standards in the design, conduct and interpretation of the trial. Often the center represents a clinical, medical or public health institution in which the clinical trial takes place. In this contribution focus is on binary outcome of the trial such as survival (yes/no), improvement of health status (yes/no), occurrence of side effect (yes/no) to mention a few of potential binary outcomes. Even if the outcome measure is continuous (such as blood pressure) it is often categorized into two possible values. Other examples include diagnostic procedures that result frequently in continuous measures. However, the outcome is almost uniquely represented in terms of result positive or result negative.

Furthermore, it is assumed that the trial is competitive in that it compares two (or more) trial arms, here denoted as *treatment* and *control* arm. A typical setting is provided in Table 1. A manifold collection of multicenter studies from various application fields is provided in the Cochrance Library (2005). In these settings, x_i^T is the number of events in the treatment arm of the i-th center, whereas n_i^T is the person-time (total of time every person

spent under risk) of the treatment in the i-th center. If all persons spend identical time under risk n_i^T is equivalent to the sample size. Analogously, for the control arm.

Interest lies in measuring the effect of treatment, frequently accomplished by means of the risk ratio $\theta = p^T/p^C$ where p^T and p^C are the risks of an event under treatment and control, respectively. Nowadays, it is widely accepted that a simple, overall estimate of the crude risk ratio estimate

$$\hat{\theta}_{crude} = \frac{\left(\sum_{i=1}^k x_i^T\right) / \left(\sum_{i=1}^k n_i^T\right)}{\left(\sum_{i=1}^k x_i^C\right) / \left(\sum_{i=1}^k n_i^C\right)}$$

is by no means a sufficient description of the available data - unless effect homogeneity is established. Mainly, two reasons are responsible for this perspective.

- The simple estimate ignores a potential center effect. In Table (1) most centers show a beneficial effect of treatment, though not always as can be seen in center 14. A more controversial example and discussion on effect heterogeneity is given in Horvitz *et al.* (1996) in which 21 centers show beneficial and 10 centers harmful effects.

- The simple estimate ignores potential covariate information. May be, age and gender distributions varied from center to center, may be, randomization failed in some centers, or may be, the background population was different from center to center.

The contribution is targeting to achieve a solid modelling of the two before mentioned situations accompanied by easy-to-use algorithms which will allow the clinician to analyse the multicenter trial in an up-to-date fashion.

4.2 Center Effect

It might be tempting to ignore the fact that study data are available for different centers. Indeed, it is possible to collapse data over all centers to achieve a simple two-by-two table from which the effect estimate could be computed simply as $\hat{\theta}_{crude} = \frac{\left(\sum_{i=1}^{k} x_i^T\right) / \left(\sum_{i=1}^{k} n_i^T\right)}{\left(\sum_{i=1}^{k} x_i^C\right) / \left(\sum_{i=1}^{k} n_i^C\right)}$.

Though this is tempting, dreadful experience has shown that calculating a *crude* risk ratio as above may lead to quite biased estimates. In fact, various confounding situations can arise: the true effect might be overestimated (inflation) or underestimated (masking), or the center might work as an effect modifier. Hence, it is advisable to take the center effect as potential confounder into account. In Table 2 the crude risk ratio is 1.74, well in the range of the center-specific risk ratios, and the center does not appear to be a confounder for this multicenter study. In another application, Arends *et al.* (2000) investigate the treatment of cholesterol lowering levels on mortality from coronary heart disease (see Table 3). Here, the crude risk ratio is 1.0770 whereas the Mantel-Haenszel adjusted risk ratio, defined as $\hat{\theta}_{MH} = \frac{\sum_{i=1}^{k} x_i^T n_i^C / n_i}{\sum_{i=1}^{k} x_i^C n_i^T / n_i}$ with $n_i = n_i^T + n_i^C$, is 0.9708, moving an elevated risk ratio to the preventive side, as it can be expected from the nature of the treatment. This example underlines the importance of considering the center-effect in all analyses.

In addition, another aspect might be worth mentioning. Whereas none of the center-specific risk ratio estimates in the Lidocaine trial confirms significantly the damaging effect of prophylactic use of Lidocaine, a center-adjusting estimator like the Mantel-Haenszel estimator will provide a significant effect $\hat{\theta}_{MH} = 1.73$ with 95% CI (1.03,2.92). Hence, it is desirable to seek optimal and valid ways to combine available information.

4.3 Sparsity

Frequently, in multicenter studies the observed data experience *sparsity*. The data are called *sparse* if the observed event counts are close to zero, occasionally in fact identical to zero. This can occur because of the event risks are very small, so that even with a large trial sparsity has to be expected. Or, the center sizes are so small (potentially because patient recruitment is extremely difficult) that even with large event risks the occurrences of low frequency counts including zero counts are likely. An example of this nature is provided in Table 4. In multicenter sparsity trials, the investigation of center-effect heterogeneity is particularly difficult, since center-specific risk ratio estimators can only be estimated with large uncertainty. In addition, the construction of a risk ratio estimator under homogeneity needs to be done with careful consideration. Here, the profile method turns out to be beneficial.

The paper is outlined as follows. In section 5, the basic model is introduced including the profile likelihood method and a discussion of it under homogeneity. In section 6, the model is extended to cope with heterogeneity. Unobserved heterogeneity is captured by means of a nonparametric mixture leading to the nonparametric mixture profile likelihood. The gradient function is introduced and the nonparametric profile maximum likelihood estimator (PNMLE) is characterized. The latter can be computed by means of the EM algorithm with gradient function update (EMGFU). This ends section 6. Section 7 provides modelling of covariate information. Elements of log-linear modelling are used and ways for finding the profile maximum likelihood estimator including their standard errors are provided. The paper ends with a discussion and putting the results into perspective.

5 The Basic Model

Modelling effect in multicenter studies is of interest and under investigation for quite some time. For an overview see Agresti and Hartzel (2000). Let x_i^T and x_i^C denote the number of events in treatment and control arm, respectively, with n_i^T the person-time in the treatment arm and n_i^C denoting the person-time in the control arm. Let the number of centers be k, so that $i = 1, ..., k$. Also, let p_i^T and p_i^C denote the risk of an event in the treatment and control arm, respectively. Typically, we will be interested in effect measures of treatment like the *risk ratio* $\theta_i = p_i^T / p_i^C$.

5.1 Likelihood

We are interested in the inference on $\theta_i = p_i^T / p_i^C$, the ratio of the two event probabilities p_i^T for the treatment arm and p_i^C for the control arm. In contrast to single study settings, in the case of a multicenter study variation of the measure of interest, here the risk ratio, between centers can be investigated. If *homogeneity* of effect can be established, the results are more *supportive* of the effect. If *heterogeneity* is present, an appropriate modelling is required and sources for it's occurrence should be investigated.

For each center and for each arm there is a Poisson Likelihood, so that for the i-th center the contribution to the likelihood of the treatment arm is

$$exp(-n_i^T p_i^T)(n_i^T p_i^T)^{x_i^T} / x_i^T! \tag{1}$$

and for the i-th center the contribution to the likelihood of the control arm

$$exp(-n_i^C p_i^C)(n_i^C p_i^C)^{x_i^C} / x_i^C! \tag{2}$$

so that that the product likelihood over *all* centers becomes

$$\prod_{i=1}^{k} exp(-n_i^T p_i^T)(n_i^T p_i^T)^{x_i^T} / x_i^T! \times exp(-n_i^C p_i^C)(n_i^C p_i^C)^{x_i^C} / x_i^C! \tag{3}$$

We find that the *log-likelihood* takes the form[1] (ignoring the only data-dependent terms)

$$\sum_{i=1}^{k} -n_i^T p_i^T + x_i^T log(p_i^T) - n_i^C p_i^C + x_i^C log(p_i^C) \tag{4}$$

5.2 Estimation of Relative Risk in Multicenter Studies

If we consider the log-likelihood (4) the question arises in which way the measure of interest like the relative risk $\theta_i = p_i^T / p_i^C$ can be entered. One can simply rewrite p_i^T as $p_i^C \theta_i$ and (4) becomes

$$\sum_{i=1}^{k} -n_i^T p_i^C \theta_i + x_i^T log(p_i^C \theta_i) - n_i^C p_i^C + x_i^C log(p_i^C) \tag{5}$$

Note that in the log-likelihood (5) occur two kinds of parameters, the effect measuring parameter θ_i and the baseline parameter p_i^C. Whereas we call the first type *parameter of interest*, the second type is called *nuisance parameter*. The nuisance parameter are not our major interest parameter, but they are necessary for a complete description of the likelihood as well as they complicate the inference.

5.3 The Profile Method in General

In more generality, let the log-likelihood $L(\mathbf{p}, \mathbf{q})$ depend on a vector $\mathbf{p}$ of parameters of interest and a vector $\mathbf{q}$ of nuisance parameters. Let $L(\mathbf{q}|\mathbf{p}) = L(\mathbf{p}, \mathbf{q})$ be the log-likelihood for arbitrary but fixed $\mathbf{p}$, and let $\mathbf{q_p}$ be such that $L(\mathbf{q_p}|\mathbf{p}) \geq L(\mathbf{q}|\mathbf{p})$ for all $\mathbf{q}$, then

$$L^*(\mathbf{p}) = L(\mathbf{q_p}|\mathbf{p}) \tag{6}$$

[1] `log` we always denote the natural logarithm, e. g. with respect to base e

is called the *profile log-likelihood*. Note that the profile log-likelihood is now depending only on the parameters of interest and, thus, the method of profile log-likelihood can be viewed as a method to deal with nuisance parameters.

5.4 The Profile Likelihood for the Effect Measure of the Risk Ratio

In this section we determine the profile log-likelihood on the basis of (5) which we now consider as a function of $\mathbf{p}^C$ for arbitrary, but fixed $\theta = (\theta_1, ..., \theta_k)'$.

$$L(\mathbf{p}^C|\theta) = \sum_{i=1}^{k} -n_i^T p_i^C \theta_i + x_i^T log(p_i^C \theta_i) - n_i^C p_i^C + x_i^C log(p_i^C) \tag{7}$$

To determine $\mathbf{p}_\theta^C$ we calculate the partial derivatives

$$\begin{aligned} \frac{\partial}{\partial p_j^C} L(\mathbf{p}^C|\theta) &= \frac{\partial}{\partial p_j^C} \sum_{i=1}^{k} -n_i^T p_i^C \theta_i + x_i^T log(p_i^C \theta_i) - n_i^C p_i^C + x_i^C log(p_i^C) \qquad (8) \\ &= -n_j^T \theta_j + x_j^T / p_j^C - n_j^C + x_j^C / p_j^C \end{aligned}$$

which can be readily solved for p_j^C as

$$p_{j\theta}^C = \frac{x_j^C + x_j^T}{n_j^C + \theta_j n_j^T}. \tag{9}$$

Inserting (9) into (7) leads to

$$\sum_{i=1}^{k} -(n_i^C + \theta_i n_i^T)(\frac{x_i^C + x_i^T}{n_i^C + \theta_i n_i^T}) + x_i^T \log(\theta_i) + (x_i^C + x_i^T) \log(\frac{x_i^C + x_i^T}{n_i^C + \theta_i n_i^T}) \tag{10}$$

which simplifies to

$$\sum_{i=1}^{k} -(x_i^C + x_i^T) + x_i^T \log(\theta_i) + (x_i^C + x_i^T) \log(x_i^C + x_i^T) + (x_i^C + x_i^T) log(n_i^C + \theta_i n_i^T) \tag{11}$$

and, finally, if we only consider parameter dependent terms

$$L^*(\theta) = \sum_{i=1}^{k} x_i^T \log(\theta_i) - (x_i^C + x_i^T) \log(n_i^C + \theta_i n_i^T). \tag{12}$$

The above expression $L^*(\theta)$ is the *profile log-likelihood* for the risk ratio and all inference will be based upon this log-likelihood.

5.5 The Profile Method under Effect Homogeneity

To illustrate the simplicity and usefulness of the profile method consider the situation of *homogeneity* of effect: $\theta_1 = \theta_2 = ... = \theta_k = \theta$. Then, taking the derivative of (12) w.r.t. θ, we yield

$$\sum_{i=1}^{k} \left(x_i^T/\theta - (x_i^C + x_i^T) n_i^T / (n_i^C + \theta n_i^T) \right) = 0, \tag{13}$$

or, equivalently

$$\sum_{i=1}^{k} w_i(\theta)/\theta \left(x_i^T n_i^C - x_i^C n_i^T \theta) \right) = 0, \tag{14}$$

where $w_i(\theta) = 1/(n_i^C + \theta n_i^T)$. Equation (14) is an implicit characterization of the maximum profile likelihood estimator of relative risk which can be further written as

$$\theta = \frac{\sum_{i=1}^{k} w_i(\theta) x_i^T n_i^C}{\sum_{i=1}^{k} w_i(\theta) x_i^C n_i^T}, \tag{15}$$

which can be used to iteratively construct the maximum likelihood estimator. If iteration is started with $\theta = 1$ it can be seen that the first iteration using (16) leads to the well-known *Mantel-Haenszel-Estimator* of the risk ratio:

$$\theta = \frac{\sum_{i=1}^{k} w_i(1) x_i^T n_i^C}{\sum_{i=1}^{k} w_i(1) x_i^C n_i^T} = \frac{\sum_{i=1}^{k} x_i^T n_i^C / n_i}{\sum_{i=1}^{k} x_i^C n_i^T / n_i}, \tag{16}$$

where $n_i = n_i^T + n_i^C$ is the total person-time of the i-th center. For the time being, we remain with the Mantel-Haenszel estimator (MHE) $\hat{\theta}_{MH} = \frac{\sum_{i=1}^{k} x_i^T n_i^C / n_i}{\sum_{i=1}^{k} x_i^C n_i^T / n_i}$ and compare it with the PMLE under effect homogeneity. Clearly, if the trial is completely balanced $n_i^T = n_i^C$ for all centers i, then the parameter-dependent weights cancel out, and PMLE and MHE are identical. Typically, MHE and PMLE are close and the loss of efficiency in using the MHE is not high. However, the exception is the situation of sparsity. Simulation studies provide some evidence that in this case there is considerable loss of efficiency, in particular, when θ is bounded away from 1. To demonstrate this finding a simulation experiment was conducted. Since differences

between PMLE and MHE can only be expected for highly unbalanced and sparse multicenter studies, n_i^T and n_i^C were generated from a Poisson with Poisson parameter 3 and 6, where arm allocation of the Poisson parameter was random to guarantee that the trial is unbalanced. The baseline parameter p_i^C was chosen uniform in [0.1,0.3] and the risk ratio parameter was kept fixed for each simulation (replication size 10,000) and risk ratio values from 0.00001 to 3.3333 were considered. The number of centers k was chosen to be 5. The results for the two estimators are provided in Fig. 1 for the bias and in Fig. 2 for the variance indicating a superior behavior of the PMLE with respect to both criteria.

In another beneficial aspect of the profile likelihood method lies in the fact that it provides easily an estimate of the variance of the PMLE. We will use a standard result from likelihood theory (see, for example Le 1992, p.72-73) that the variance of the maximum likelihood estimate can be approximated by negative inverse of the second derivative of the log-likelihood function which is evaluated at the maximum likelihood estimate. We apply this result to the profile likelihood situation. Let us write the profile log-likelihood (12) using $\phi = log(\theta)$

$$L^*(\phi) = [\sum_{i=1}^{k} x_i^T]\phi - \sum_{i=1}^{k}(x_i^C + x_i^T)log(n_i^C + e^{\phi} n_i^T). \tag{17}$$

with second derivative

$$L^{*''}(\phi) = -\sum_{i=1}^{k} \frac{x_i n_i^T n_i^C e^{\phi}}{(n_i^C + e^{\phi} n_i^T)^2} = -\sum_{i=1}^{k} \frac{x_i n_i^T n_i^C \theta}{(n_i^C + \theta n_i^T)^2} = -\sum_{i=1}^{k} x_i \alpha_i (1-\alpha_i), \tag{18}$$

so that an estimate of the variance of the PMLE of ϕ is provided as

$$\widehat{var(\hat{\phi})} = \widehat{var(\log \hat{\theta})} = \left(\sum_{i=1}^{k} x_i \alpha_i (1-\alpha_i) \right)^{-1}, \tag{19}$$

where $\alpha_i = \frac{\hat{\theta} n_i^T}{n_i^C + \hat{\theta} n_i^T}$. The Mantel-Haenszel estimate of the common relative risk is given as $\hat{\theta}_{MH} = \frac{\sum_i x_i^T n_i^C / n_i}{\sum_i x_i^C n_i^C / n_i}$ with $n_i = n_i^C + n_i^T$. Although the formula

for Mantel-Haenszel relative risk estimate is quite elementary, a widely accepted expression for its variance has been only given recently (Greenland and Robins 1985; see also Woodward 1999, p. 170). We have that

$$\widehat{var}(log(\hat{\theta}_{MH})) = \frac{\sum_i (n_i^T n_i^C x_i - x_i^T x_i^C n_i)/(n_i)^2}{(\sum_i x_i^T n_i^C/n_i)(\sum_i x_i^C n_i^T/n_i)} \tag{20}$$

where $x_i = x_i^C + x_i^T$ as before. We note that (20) has been developed for the situation of identical person-times in the centers reflecting a binomial sampling plan. Breslow (1984) provided a robust variance formula for the situation of person-specific observation times.

Typically, not only MHE and PMLE, but also the variances of the MHE (20) and of the PMLE (19) are close. However, the exception is the situation of sparsity. Simulation studies provide some evidence that in this case the variance estimator (19) is behaving better than (20). To demonstrate this fact the following simulation experiment was conducted. In the balanced trial PMLE and MHE are identical, so that direct comparability of (19) and (20) is available. A sparse balanced multicenter trial was simulated, with $n_i^T = n_i^C$ being generated from a Poisson with Poisson parameter 5. The baseline parameter p_i^C was chosen uniform in [0.1,0.3] and the risk ratio parameter was kept fixed for each simulation (replication size 10,000) and risk ratio values from 0.00001 to 3.3333 were considered. The number of centers k was chosen to be 20. The results for the two variance estimators are provided in Fig. 3 for the bias of the variance estimators and in Fig. 4 for the variance of the variance estimators indicating a slightly better behavior for (19) with respect to both criteria.

6 Modelling Unobserved Heterogeneity

6.1 Unobserved Covariate and the Marginal Profile Likelihood

In this section, a general approach for coping with center-effect heterogeneity is proposed. Assume that the population of all centers consists out of m subpopulations with weights q_j and subpopulation risk ratio θ_j. Let us consider again the likelihood (3) where we - for simplicity of presentation - consider only a single center:

$$Po(x^T, n^T p^C \theta) \times Po(x^T, n^T p^C) \tag{21}$$

which becomes - after replacing p^C by their conditional maximum likelihood estimates $\frac{x^C+x^T}{n^C+\theta n^T}$

$$Po(x^T, n^T \frac{x^C + x^T}{n^C + \theta n^T}\theta) \times Po(x^C, n^C \frac{x^C + x^T}{n^C + \theta n^T}), \tag{22}$$

where $Po(x, \lambda) = \exp(-\lambda)\lambda^x/x!$. Consider next the situation that for each observation $(x^T, n^T, x^C, n^C)'$ there is an unobserved m-vector $\mathbf{y}$ with a 1 in the $j-$th position (and 0 otherwise) assigning the component population j to which the observation belongs to. Taking the margin over the unobserved vector $\mathbf{y}$ leads to the *marginal density*

$$\begin{array}{c} \sum_{\mathbf{y}} Po(x^T, n^T \frac{x^C+x^T}{n^C+\theta_y n^T}\theta_y) \times Po(x^C, n^C \frac{x^C+x^T}{n^C+\theta_y n^T})q_y \\ = \sum_{j=1}^m Po(x^T, n^T \frac{x^C+x^T}{n^C+\theta n^T}\theta_j) \times Po(x^C, n^C \frac{x^C+x^T}{n^C+\theta_j n^T})q_j \end{array} \tag{23}$$

where $q_y = q_j$ is the weight which the $j-$th population with parameter value $\theta_y = \theta_j$ receives. Taking now the log-likelihood over all centers and ignoring terms that do not involve the parameters the following *mixture profile log-*

likelihood is achieved:

$$\begin{aligned}\sum_{i=1}^{k}\log\left[\sum_{j=1}^{m}\exp\left(-\tfrac{x_i n_i^T\theta_j}{n_i^C+\theta_j n_i^T}\right)\theta_j^{x_i^T}\times\exp\left(-\tfrac{x_i n_i^C}{n_i^C+\theta_j n_i^T}\right)\left(\tfrac{1}{n_i^C+\theta_j n_i^T}\right)^{x_i}q_j\right]\\ =\sum_{i=1}^{k}\log\left[\sum_{j=1}^{m}\exp(-x_i)\;\theta_j^{x_i^T}\left(\tfrac{1}{n_i^C+\theta_j n_i^T}\right)^{x_i}q_j\right]\\ \propto\sum_{i=1}^{k}\log\left[\sum_{j=1}^{m}\theta_j^{x_i^T}\left(\tfrac{1}{n_i^C+\theta_j n_i^T}\right)^{x_i}q_j\right]=L^*(Q)\end{aligned}\tag{24}$$

which we may sometimes write as

$$L^*(Q)=\sum_{i=1}^{k}\log\left(\sum_{j=1}^{m}f_i(\theta_j)q_j\right)\tag{25}$$

where $f_i(\theta_j)=\frac{\theta_j^{x_i^T}}{(n_i^C+\theta_j n_i^T)^{x_i}}$ and $x_i=x_i^C+x_i^T$. Also, Q denotes the discrete probability distribution $Q=\begin{pmatrix}\theta_1 & \dots & \theta_m\\ q_1 & \dots & q_m\end{pmatrix}$ giving mass q_j to the risk ratio θ_j in subpopulation j, also called the *mixing distribution*.

6.2 Concavity, the Gradient Function and the Mixture Maximum Likelihood Theorem

6.1 It is easy to verify that $L^*(Q)$ is a *concave* functional in the set Ω of all discrete probability distributions, though this is not necessarily the case for Ω_m, the set of all discrete probability distributions with *exactly* m support points (subpopulations). Hence, a global *profile mixture maximum likelihood estimator* (PNMLE) exists, but the number of support points is itself part of the estimation process. Let us define the *gradient function* as an important tool for finding the PNMLE. In particular, let for arbitrary but fixed $Q=\begin{pmatrix}\theta_1 & \dots & \theta_m\\ q_1 & \dots & q_m\end{pmatrix}$ and any $\theta>0$

$$d(\theta,Q)=\frac{1}{k}\sum_{i=1}^{k}\frac{f_i(\theta)}{\sum_{j=1}^{m}f_i(\theta_j)q_j}\tag{26}$$

The gradient function (26) can be motivated by means of the concept of the directional derivative where it is contained as the essential part (for details

see Lindsay 1983, Lindsay 1995, Böhning 2000). A first major application arises in the *general mixture maximum likelihood theorem* which states that $\hat{Q} = \begin{pmatrix} \hat{\theta}_1 & \dots & \hat{\theta}_m \\ \hat{q}_1 & \dots & \hat{q}_m \end{pmatrix}$ is PNMLE if and only if $d(\theta, \hat{Q}) \leq 1$ for all $\theta > 0$. In addition, for the support points of $\hat{Q}$ we have that the upper bound becomes sharp, e.g. $d(\hat{\theta}_j, \hat{Q}) = 1$. As a first consequence, we might be able to identify effect homogeneity without further testing whatsoever. Indeed, let $\hat{\theta}_{PMLE}$ denote the profile maximum likelihood estimator under homogeneity. If

$$d(\theta, \hat{\theta}_{PMLE}) = \frac{1}{k} \sum_{i=1}^{k} \frac{f_i(\theta)}{f_i(\hat{\theta}_{PMLE})} \leq 1,$$

for all $\theta > 0$, then $\hat{\theta}_{PMLE}$ must be the PNMLE, and no further search for heterogeneity is necessary.

Lidocaine Trial. We come back to the multicenter study presented in Table 2. A graph of the gradient function $\theta \rightarrow d(\theta, \hat{\theta}_{PMLE})$ for the maximum likelihood estimator $\hat{\theta}_{PMLE}$ of θ under homogeneity (more precisely, the one-point probability measure giving all mass to $\hat{\theta}_{PMLE}$) is provided in Fig. 5 showing clear evidence of homogeneity, making further testing for heterogeneity unnecessary.

Cholesterol Lowering Treatment and Coronary Heart Disease. Let us consider again the multicenter study presented in Table 3. Here, the graph (see Fig. 6) of the gradient function $\theta \rightarrow d(\theta, \hat{\theta}_{PMLE})$ for the maximum likelihood estimator $\hat{\theta}_{PMLE}$ of θ under homogeneity indicates clear evidence of heterogeneity. The upper bound one is violated (see Fig. 6) and $\hat{\theta}_{PMLE}$ can not be the PNMLE. In fact, it is clear that the PNMLE will have more than one support point.

For the general construction of the PNMLE, numerical algorithms will be required.

6.3 The Nonparametric Profile Maximum Likelihood Estimator via the EM Algorithm

A major tool for constructing the maximum likelihood estimates is the EM algorithm (Dempster, Laird and Rubin 1977, McLachlan and Krishnan 1996). It requires the specification of a suitable *complete data likelihood* which for mixtures is conventionally taken as

$$\prod_{i=1}^{k}\prod_{j=1}^{m}\left(Po(x_i^T, n_i^T\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T}\theta_j)\times Po(x_i^C, n_i^C\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T})\times q_j\right)^{y_{ij}}, \quad (27)$$

where $y_{ij}=1$, if center i belongs to subpopulation j, and 0 otherwise. Since y_{ij} is unobserved, it is replaced in the *E-step* of the EM-algorithm by their expected values

$$\begin{aligned} e_{ij} &= E(Y_{ij}|Q,\ \text{data}\ ) \\ &= \frac{Po(x_i^T, n_i^T\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T}\theta_j)\times Po(x_i^C, n_i^C\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T})\times q_j}{\sum_{j'=1}^{m} Po(x_i^T, n_i^T\frac{x_i^C+x_i^T}{n_i^C+\theta_{j'} n_i^T}\theta_{j'})\times Po(x_i^C, n_i^C\frac{x_i^C+x_i^T}{n_i^C+\theta_{j'} n_i^T})\times q_{j'}}. \end{aligned} \quad (28)$$

Replacing y_{ij} in (27) by their expected values leads to the *expected* complete data likelihood

$$\prod_{i=1}^{k}\prod_{j=1}^{m}\left(Po(x_i^T, n_i^T\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T}\theta_j)\times Po(x_i^C, n_i^C\frac{x_i^C+x_i^T}{n_i^C+\theta_j n_i^T})\times q_j\right)^{e_{ij}}, \quad (29)$$

which can be maximized in θ_j and q_j, separately. This established the *M-step* of the EM algorithm. In fact, we find easily that

$$\hat{q}_j = \frac{1}{k}\sum_{i=1}^{k} e_{ij}.$$

Furthermore, $\hat{\theta}_j$ can be found from the equation

$$\sum_{i=1}^{k}\frac{e_{ij}x_i^T}{\theta_j} - \frac{e_{ij}x_i n_i^T}{n_i^C+\theta_j n_i^T} = 0$$

by using the iteration

$$\hat{\theta}_j = \frac{\sum_{i=1}^{k} e_{ij} x_i^T n_i^C w_i(\hat{\theta}_j)}{\sum_{i=1}^{k} e_{ij} x_i^C n_i^T w_i(\hat{\theta}_j)} \tag{30}$$

with $w_i(\theta) = 1/(n_i^T\theta + n_i^C)$, in analogy to the homogenous case (15).

6.4 The EMGFU for the Profile-Likelihood-Mixture

When the gradient function indicates heterogeneity, usually the number of components adequate to model this heterogeneity will be unknown and several values for m need to be considered. Hence, it appears appropriate to consider *all* possible values of m, starting from $m = 1$ to the number of components involved in the PNMLE. The following algorithm is in analogy to the *EM algorithm with gradient function update* (Böhning 2003).

The initial step starts with the case of homogeneity ($m = 1$) and the computation of the profile maximum likelihood estimator under homogeneity. If the gradient function violates the upper bound, e.g. $d(\theta_{max}, \theta_{PMLE}) > 1$, then the number of components is increased to $m = 2$ and the EM algorithm of the previous section is utilized with initial values for the two components $\theta_1 = \theta_{PMLE}$ and $\theta_2 = \theta_{max}$ to compute a discrete two-support size probability distribution $Q^{(2)}$. Otherwise (if the gradient function does not violate the upper bound), the algorithm is stopped.

Now suppose that the EM algorithm has generated for current value of m a discrete probability distribution $Q^{(m)}$ having support points $\theta_1, ..., \theta_m$. If the gradient function violates the upper bound, e.g. $d(\theta_{max}, Q^{(m)}) > 1$, then the number of components is increased to $m = m + 1$ and the EM algorithm of the previous section is utilized with initial values for the $m + 1$ components $\theta_1, ..., \theta_m$ and $\theta_{m+1} = \theta_{max}$ to compute a discrete $(m + 1)$-support size probability distribution $Q^{(m+1)}$. Otherwise (if the gradient function does not violate the upper bound), the algorithm is stopped. This step is repeated until the PNMLE is reached. The advantage of the EMGFU lies in the fact

that it combines a strategy for generating the nonparametric profile maximum likelihood estimator with a search for the best local mixture maximum likelihood estimator with exactly m components.

Cholesterol Lowering Treatment and Coronary Heart Disease. We would like to demonstrate the EMGFU for the multicenter trial given in Table 3. In this case, we had found clear evidence of heterogeneity (see section 6.1). Table 5 provides details on this analysis. The EMGFU algorithm starts with homogeneity and provides $\theta_{PMLE} = 0.9716$, then increases the number of support points stepwise by means of the gradient function until the nonparametric profile maximum likelihood estimator with $m = 4$ components is reached. Indeed, from gradient function plot (Figure 7) it is evident that the global nonparametric profile maximum likelihood estimator has been reached.

6.5 Likelihood Ratio Testing, Model Evaluation, and Classification of Centers

Profile likelihoods behave similar to likelihoods. However, for mixture models this just means that we have to face the same problems. Profile likelihood ratios will not have standard χ^2-distributions, so that choices for the number of components, solely based upon the likelihood ratio, might be misleading and should be accompanied by other selection criteria such as the *Bayesian Information Criterion* which has proved to be valuable selection criterion in other settings. McLachlan and Peel (2000) discuss and compare various selection criteria. Within the simpler criteria, the *Akaike Information Criteria* shows a tendency to select too many components (overestimate m), whereas the BIC, though not always correct, behaves better. In Table 5 can be seen that the best model according to BIC is the model with two components.

Note that for i fixed, e_{ij}, as given in (28), is a probability distribution. In fact, e_{ij} is the posterior probability for center i belonging to subpopulation j.

This enables to classify center i into that subpopulation j where the posterior probability is the largest. Recall that for the cholesterol lowering trial there was a PNMLE found consisting out 4 subpopulations. However, according to the BIC, only 2 subpopulations are required. Table 6 provides the posterior probabilities for each of the 33 centers including a classification of each center into the component associated with the highest posterior.

7 Modelling Covariate Information

7.1 A Multicenter Trial with Covariate Information

Frequently, a multicenter trial does not only provide information on treatment and control arm, outcome and sample sizes, but also include further, potentially quite important co-information, which might show some joint variation with the effect of interest. In the following such an example is provided.

DuMouchel and Normand (2000) discuss a multicenter study with 59 centers on smoking cessation (see Figure 3.1). Here, the event of interest is quit smoking and treatment varied from trial to trial (patch versus gum and low versus high intensity support). It might be of interest to see if the effect (the success risk ratio of quitting smoking) is depending on the kind of support (low/high) and on the kind of device (gum/patch) used. In addition, the quit rate might be affected by the time of study (general time trends). Therefore, one might be interested in investigating two binary covariate effects(gum/patch and low/high support) and one of continuous nature (time) on the success relative risk of quitting smoking.

7.2 A Generalized Linear Model

It is assumed that the co-information is available in terms of covariates $z_1, z_2, ..., z_p$ which can be put into a vector $\mathbf{z} = (z_1, ..., z_p)'$. z_1 might be

the proportion of women in the center, z_2 the mean age in the center, among other potential co-information. For *center* this vector has a certain value $\mathbf{z}_i$ where then z_{i1} denotes the proportion of women in center i (in the hypothetical example), z_{i2} denotes the mean age in center i, and so forth. We consider again (12):

$$L^*(\theta) = \sum_{i=1}^{k} x_i^T \log(\theta_i) - (x_i^C + x_i^T) \log(n_i^C + \theta_i n_i^T). \tag{31}$$

How can the covariate information be linked to the relative risk parameter θ_i? This can be done by means of an appropriate modification of the idea of a generalized linear model (McCullagh and Nelder 1989). Here, the *linear predictor* $\eta_i = \beta' \mathbf{z}_i = \beta_1 z_{i1} + \beta_2 z_{i2} + ... + \beta_p z_{ip}$ needs to be linked to the relative risk parameter θ_i. Typically, $z_{i1} = 1$, so that the model contains an intercept parameter. From basic regression courses the *identity link* comes up into mind: $\theta_i = \eta_i = \beta' \mathbf{z}_i$. However, though simple, it is not appropriate since it does not guarantee that $\theta_i \geq 0$ which would be an essential requirement for a relative risk. One good choice of a link is the *canonical link* $\theta_i = \exp(\eta_i)$ which guarantees $\theta_i \geq 0$. To illustrate the model suppose that we are describing the covariates of section 4.1 for the smoking cessation trials. z_1 might be the binary covariate describing if a patch ($z_1 = 1$) was used or gum ($z_1 = 0$) and z_2 might be the binary covariate describing if low ($z_2 = 0$) or high ($z_2 = 1$) support was provided. Then, the linear predictor $\eta_i = \beta_0 + \beta_1 z_1 + \beta_2 z_2$ can capture all situations easily. For example, the combination "gum" and "low support" leads to $\eta_i = \beta_0$, or the combination "patch" and "high support" leads to $\eta_i = \beta_0 + \beta_1 + \beta_2$, assuming that there is no interaction. Of course, we need to establish if any of the effects are significant. For this we need the likelihood.

Under the *log-link* the associated likelihood simplies to

$$L^*(\beta) = \sum_{i=1}^{k} x_i^T \eta_i - (x_i^C + x_i^T) \log(n_i^C + \exp(\eta_i) n_i^T), \tag{32}$$

with $\eta_i = \beta' \mathbf{z}_i$. The log-likelihood (32) needs to be maximized in β.

7.3 Finding Maximum Profile Likelihood Estimates

For finding the maximum likelihood estimators we need to maximize the log-likelihood (32). For this purpose consider the partial derivative w.r.t. β_j

$$\frac{\partial L^*}{\partial \beta_j}(\beta) = \sum_{i=1}^{k} x_i^T z_{ij} - x_i n_i^T \frac{\exp(\eta_i)}{n_i^C + \exp(\eta_i) n_i^T} z_{ij}$$

and the corresponding vector of partial derivates, the *gradient*:

$$\nabla L^*(\beta) = (\frac{\partial L^*}{\partial \beta_1}, ..., \frac{\partial L^*}{\partial \beta_p})'. \tag{33}$$

Furthermore, the *Hesse* matrix of second derivatives is

$$\frac{\partial^2 L^*}{\partial \beta_h \partial \beta_j}(\beta) = -\sum_{i=1}^{k} \frac{x_i n_i^T n_i^C \exp(\eta_i)}{(n_i^C + \exp(\eta_i) n_i^T)^2} z_{ij} z_{ih} \tag{34}$$

so that (34) becomes in matrix form

$$\nabla^2 L^*(\beta) = \left(\frac{\partial^2 L^*}{\partial \beta_h \partial \beta_j}(\beta) \right) = -\mathbf{Z}' \mathbf{W}(\beta) \mathbf{Z} \tag{35}$$

where $\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & ... & z_{1p} \\ z_{21} & z_{22} & ... & z_{2p} \\ . & . & ... & . \\ z_{k1} & z_{k2} & ... & z_{kp} \end{pmatrix}$ is the *design matrix*, independent of β and a diagonal matrix $\mathbf{W}(\beta) = (w_{ij}(\beta))$, with $w_{ij} = 0$, if $i \neq j$, and $w_{ii} = \frac{x_i n_i^T n_i^C \exp(\eta_i)}{(n_i^C + \exp(\eta_i) n_i^T)^2}$.

The Newton-Raphson procedure for iteratively constructing the maximum likelihood estimates can now easily be given. Choose some $\beta^{(0)}$ as initial value (for example $\beta^{(0)} = 0$) and then update β according to

$$\beta^{(n+1)} = \beta^{(n)} - \nabla^2 L^*(\beta^{(n)})^{-1} \nabla L^*(\beta^{(n)}) \tag{36}$$

until convergence. The convergence of the sequence (36) is not guaranteed. Since in this case $w_{ii} = \frac{x_i n_i^T n_i^C \exp(\eta_i)}{(n_i^C + \exp(\eta_i) n_i^T)^2} \leq x_i/4$ for all values of η_i, it is possible

to replace the Newton-Raphson step by the lower bound procedure (Böhning and Lindsay 1988, Lange 2004), leading to

$$\beta^{(n+1)} = \beta^{(n)} + \mathbf{B}^{-1}\nabla L^*(\beta^{(n)}) \tag{37}$$

where $\mathbf{B} = \mathbf{Z}'\Lambda\mathbf{Z}$ and Λ is a diagonal matrix with $\Lambda_{ii} = x_i/4$, independent of η. $\mathbf{B}$ represents a global upper bound for $-\nabla^2 L^*(\beta)$, e.g. $-\nabla^2 L^*(\beta) \leq \mathbf{B}$ for all β, where "$\leq$" denotes the matrix ordering. Besides it's guaranteed convergence to the maximum, (37) has the advantage to have the global bound matrix be inverted only once (for details see Böhning and Lindsay 1988, Böhning 1992). Note that in ordinary log-linear (Poisson) regression the second-derivative matrix is unbounded and the strong result of global convergence, as in the case here, can not be achieved.

7.4 Finding Standard Errors

Estimated variances of maximum likelihood estimates can be found from the negative inverse of the *information matrix* (35), namely $\left(\mathbf{Z}'\mathbf{W}(\hat{\beta})\mathbf{Z}\right)^{-1}$, in particular, $\widehat{var}(\hat{\beta}_j) = \left(\mathbf{Z}'\mathbf{W}(\hat{\beta})\mathbf{Z}\right)^{-1}_{jj}$, where $\hat{\beta}$ is the vector of maximum likelihood estimates. These variance estimates are obtained as a by-product of the Newton-Raphson iteration (36). Significance of individual effects can be consequently obtained by means of a Wald-test (or similar procedures) using

$$T_j = \frac{\hat{\beta}_j}{\widehat{s.e.}(\hat{\beta}_j)},$$

where $\widehat{s.e.}(\hat{\beta}_j) = \sqrt{\widehat{var}(\hat{\beta}_j)}$. Here, it would be argued that - under the null-hypothesis of no effect of the j-th covariate - T_j is asymptotically normally distributed.

Smoking cessation and treatment variation. We come back to the multicenter trial on the evaluation of the success in quitting smoking. The

results of fitting various models are provided in Table 7. Evidently, the treatment modification of using a patch (vs. gum) yields the only significant effect change for quitting smoking.

8 Comparison with Other Approaches

8.1 Approximate Likelihood

Besides the profile approach there are other methods to estimate the treatment effect in a MAIPD. A conventional approach assumes a normal distribution for the logarithmic relative risk in which for each center the variance is gained from a first-order Taylor-series approximation. For the analysis, this variance is treated as a known value. Modelling of unobserved heterogeneity has been discussed Laird (1987) and DerSimonian and Laird (1986). Modelling covariate information for this situation has been discussed in Hedges (1994) and, in more generality, in Hedges and Cooper (1994), Thompson and Sharp (1999), Brockwell and Gordon (2001) and Böhning (2000) among others. A general introduction is also given by Houwelingen *et al.* (2000). The problem with this approach is, for one, the potentially insufficient approximation of the normal distribution, and, for two, that the trial-specific variances are treated as known and fixed values (see also the discussion on this issue in Böhning *et al.* 2002), and, if estimated, the variance estimates might be poor, especially if the event frequencies are small.

8.2 Multi-Level Approach

An approach, which is overcoming these problems, is the hierarchical, multilevel approach (see for example Turner *et al.* 2000, Goldstein 1995). For the situation here, the two-level hierarchical model has been outlined clearly in Aitkin (1999a,b) including illuminating demonstrations at various MAIPDs (see also Aitkin and Alfó 1998). The approach utilizes exact likelihoods and

does not rely on approximations. For comparison, let us consider this approach in a little more detail. The first level is modelled by means of a log-linear model with $\log(p_i^C) = \alpha_i$ and $\log(p_i^T) = \alpha_i + \beta_i$ where p_i^C and p_i^T are as defined previously. Clearly, α_i represents in this approach the baseline parameter and β_i is the log-risk ratio. Under the Poisson assumption the likelihood for study i is given as $Po(x_i^T, p_i^T n_i^T) \times Po(x_i^C, p_i^C n_i^C)$. If the parameters p_i^C and p_i^T are replaced by their log-linear reparameterizations $p_i^C = e^{\alpha_i}$ and $p_i^T = e^{\alpha_i+\beta_i}$ the following likelihood for the i-th study is provided:

$$f_{ML}(x_i^T, x_i^C|\alpha_i, \beta_i, n_i^T, n_i^C) = Po(x_i^T, e^{\alpha_i+\beta_i} n_i^T) \times Po(x_i^C, e^{\alpha_i} n_i^C) \qquad (38)$$

The second level will be modelled by means of a non-parametric mixing distribution. We consider the most complex form of heterogeneity, allowing for baseline *and* effect heterogeneity, that is, each component in the mixture has its own baseline and effect parameter. Accordingly, the mixture distribution has the form:

$$f(x_i^T, x_i^C|Q^*, n_i^T, n_i^C) = \sum_{j=1}^{m} f_{ML}(x_i^T, x_i^C|\alpha_j, \beta_j, n_i^T, n_i^C) q_j$$

where the mixing distribution Q^* is now given as:

$$Q^* = \begin{pmatrix} \alpha_1 & \ldots & \alpha_m \\ \beta_1 & \ldots & \beta_m \\ q_1 & \ldots & q_m \end{pmatrix}.$$

Results of Modelling for the Cholesterol Lowering MAIPD. Let us illustrate the approach with the MAIPD on cholesterol lowering treatment in Table 3. The nonparametric MLE has here 15 components (see Table 8). In contrast, the PNMLE has 4 components (see Table 8).

Table 8 about here

As can be seen in Table 8 more components are estimated by the nonparametric maximum likelihood estimator in the multi-level model in comparsion to the profile likelihood approach (15 vs. 4). According to the BIC the multi-level approach provides 7 components, whereas the profile likelihood approach suggests only 2 components (see Table 5). One reason for this behavior becomes evident by inspecting the nonparametric MLE (or the one with 7 components returned from the BIC) more closely. The multi-level model experiences increased heterogeneity through the baseline parameter (which has low practical interest) and many components have close treatment effect estimates.

8.3 Some simulation results

To investigate this confounding baseline effect in the multi-level model a bit further by means of a small simulation study, it is assumed that in the population of interest two treatment clusters exist. The clusters are represented by the mixing distribution $Q = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$. In other words, both components, one with risk ratio 0.5 and the other with risk ratio 1.5 receive the same weight of 0.5. To mimic baseline variation the baseline risk $p_1^C, ...p_k^C$ were generated from an uniform distribution on 0.1 to 0.66. The parameter $p_i^T = \theta_j p_i^C$ is then provided from a Bernoulli experiment with event probability 0.5 for $j = 1$ or $j = 2$. The sample sizes n_i^T and n_i^C were generated from a Poisson distribution with parameter 100. Then, Poisson variates X_i^T with parameters $n_i^T \times p_i^T$ and Poisson variates X_i^C with parameters $n_i^C \times p_i^C$ were drawn for each study i, $i = 1, ..., k$. In this case the number of studies was $k = 100$. To achieve better comparibility, for both, the profile likelihood and the multi-level model, a two component mixing distribution were estimated. The procedure was replicated 1,000 times. From the resulting replications mean and variance of each component effect parameter estimate

were computed. The result for this constellation is visualized in Figure 9 (upper panel).

Figure 9 about here

The first component parameter is considerably overestimated in the multi level model. Note that actually the true relative risk is *not* captured by the 95% confidence interval. In contrast the second component is likewise underestimated. The pofile likelihood behaves clearly superior in finding the true mixing distribution. To show the large influence of the baseline parameter we have provided in Figure 9 (lower panel) a further result from the simulation study. The simulation setting is as in the previous one with the exception that we now assume a fixed baseline parameter $p_i^C = 0.3$. Now both models are capable in recovering the true mixing distribution and have a confidence intervals of comparable size and location.

A final aspect of a wide scope of simulation results concerns the effect homogeneity case. The main settings are: $p_1^C, ...p_k^C$ were generated from an uniform distribution on 0.3 to 0.6; $p_i^T = \theta p_i^C$ and θ is fixed for all studies $i = 1, ..., k$; n_i^T and n_i^C were generated from a Poisson distribution with parameter 100; the number of studies is $k = 100$; X_i^T with parameters $n_i^T \times p_i^T$ and X_i^C with parameters $n_i^C \times p_i^C$ were generated from an Poisson

Figure 10 about here

distribution. In Figure 10 the bias is provided for both models for 30 values of θ ranging from 0.1 to 1.32. The bias is acceptably small for both models (as are the variances close for both estimators), although the profile likelihood estimator appears to be virtually unbiased for a wide range of θ-values.

9 Other Issues Investigated

9.1 Theoretical Aspects of the Profile Likelihood Methodology

The profile method is a traditional method of dealing with nuisance parameters. One of it's critical aspects centers on a potential overprecision in the estimator for the parameter of interest "resulting from apparently knowing the nuisance parameter as an explicit function of the data and the parameter of interest" (Aitkin 1998) has led to several proposals of adjusting or modifying it (Barndorff-Nielsen and Cox 1994). On the other hand, for a finite parameter, the curvature based on the profile likelihood is identical to the curvature achieved from the parameter-of-interest part in the full likelihood curvature (Patefield 1977). But even for semi-parametric settings, as discussed in Murphy and Van der Vaart (2000) under mild regularity assumptions, profile likelihoods behave like ordinary likelihoods. In particular, variance approximations can be found from utilizing second derivatives of the profile log-likelihood in the conventional way.

In the ideal case (Pawitan 2001), parameter of interest and nuisance parameter are orthogonal, that is, the joint likelihood $\pounds(\theta, p^C) = \pounds_1(\theta)\pounds_2(p^C)$ factors into likelihood depending only on θ and p^C, respectively. For the ease of discussion only one trial is considered, though generalizations are straightforward. Write the joint likelihood $\exp(-p^T n^T)(p^T n^T)^{x^T} \times \exp(-p^C n^C)(p^C n^T)^{x^T}$ as product of $\pounds_1(\theta) = \left(\frac{n^T\theta}{n^C+n^T\theta}\right)^{x^T}\left(\frac{n^C}{n^C+n^T\theta}\right)^{x^C}$ and $\pounds_2(\eta_n) = \exp(-\eta_n)\eta_n^{x^T+x^C}$, where θ is the risk ratio and $\eta_n = n^T p^T + n^C p^C$. In case that the trial is balanced $\eta_n = n^T p^T + n^C p^C = \eta(n^T + n^C)$, and θ and $\eta = p^T + p^C$ 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 $\eta_n = n^T p^T + n^C p^C$ neccessarily incorporates the known, trial-specific sample size parameters,

but $\mathcal{L}_1(\theta)$ will remain identical. Alternatively, one may base inference on the likelihood conditional on the sufficient statistic $x = x^T + x^C$ for the nuisance parameter, and, although this is by no means in generality the case, it does turn out again to be $\mathcal{L}_1(\eta)$ (see for a more general discussion Pawitan 2001 or McCullagh and Nelder 1989).

In our context, assuming a finite parameter would mean that the number of centers is considered fixed, and the asymptotics refer to the number of persons within each center. But even for semi-parametric settings, as discussed in Murphy and Van der Vaart (2000) under mild regularity assumptions, profile likelihoods behave like ordinary likelihoods. In particular, variance approximations can be found from utilizing second derivatives of the profile log-likelihood in the conventional way. This theoretical fact confirms our simulation results from section 5.

9.2 Software Developments: C.A.MAP

The log-likelihood (32) is non-standard, it is *not* a Poisson log-likelihood, nor any of the log-likelihoods available in the standard generalized linear model family correspond to it. This makes it less attractive to use one of existing statistical packages like `STATA`, `S-plus`, `MINITAB`, or any package able to do macro-like programming. In addition, the available global bound (with respect to the matrix ordering) for the second derivative matrix will allow to use more reliable algorithms for computing profile maximum likelihood estimators. Hence, a software tool was developed, Computer-Assisted Analysis of Multi-Level Model, Approximate, and Profile Likelihood by Mixtures `(C.A.MAP)` which can handle the tasks of computing and inference for the models developed in the previous chapters. All examples have been analyzed with this software.

9.3 Incorporating Covariate Information and Unobserved Heterogeneity

One of the remaining tasks is to extend the modelling to investigate for unobserved heterogeneity in the presence of observed covariates. The log-likelihood corresponding to (32) and adjusting for unobserved heterogeneity is

$$L^*(\beta, Q) = \sum_{i=1}^{k} \log \left(\sum_{j=1}^{m} q_j \frac{\exp(\beta_0^{(j)} + \eta_i)^{x_i^T}}{(n_i^C + \exp(\beta_0^{(j)} + \eta_i) n_i^T)^{x_i}} \right), \qquad (39)$$

where Q is a discrete distribution giving weights $q_1, q_2, ..., q_m$ to $\beta_0^{(1)}, \beta_0^{(2)}, ..., \beta_0^{(m)}$ in the linear predictor $\beta_0^{(j)} + \eta_i = \beta_0^{(j)} + \beta_1 z_{i1} + \beta_2 z_{i2} + ... + \beta_p z_{ip}$. In other words, it is allowed that an unobserved covariate can enter *linearly* into the model. However, it is one of the complications that *residual heterogeneity* can enter in numerous ways and (39) illuminates only one of many ways. The intercept might be fixed and mixing might occur in one, several or all covariate parameters, as it might occur in the intercept parameter as well. In other words, a new range of models need to be considered and this will be approached in forthcoming work.

References

1. Agresti, A. and Hartzel, J. (2000). Strategies for comparing treatments on binary response with multicenter data. Tutorial in Biostatistics. *Statistics in Medicine*, **19**, 1115-1139.

2. Aitkin, M. (1998). Profile Likelihood. In: *Encyclopedia of Biostatistics*, Volume 5, 3534-3536, Wiley, Chichester, New York.

3. Aitkin, M. (1999a). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117-128.

4. Aitkin, M. (1999b). Meta-analysis by random effect modelling in generalized linear models. *Statistics in Medicine* **18**, 2343-51.

5. Aitkin, M., Alfó, M. (1998). Regression models for binary longitudinal responses. *Statistics and Computing* **8**, 289-307.

6. Arends, L.R., Hoes, A.W., Luben, J., Grobbee, D.E., Stijnen, T. (2000). Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* **19**, 3497-3518.

7. Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman & Hall, London.

8. Böhning, D. & Lindsay, B.-G. (1988). Monotonicity of quadratic approximation algorithms. *Annals of the Institute of Statistical Mathematics,* **40**, 223-244.

9. Böhning, D. (1992). Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics* **44**, 197-200.

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

11. Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P. (2002). Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. *Biostatistics* **3**, 445-457.

12. Böhning, D. (2003). The EM Algorithm with gradient function update for discrete mixtures with known (fixed) number of components. *Statistics and Computing,* **13**, 257-265.

13. Breslow N.E. (1984). Elementary methods of cohort analysis. *International Journal of Epidemiology* **13**, 112-115.

14. Brockmell, S.E., Gordon, I.R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine* **20**, 825-840.

15. Bryant, J., Fisher, B., Gunduz, N., Costantino, J.P., Emir, B. (1998). S-Phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, Estrogen-Receptor Positive Breast Cancer. *Breast Cancer Research and Treatment*, **51**, 239-253.

16. The Cochrane Library (2005). The cochrane collaboration, http://-www3.interscience.wiley.com/aboutus/sharedfiles/cochrane_transition.

17. Cooper, M.R., Dear, K.B.G., McIntyre, O.R., Ozer, H., Ellerton, J., Cannellos, G., Duggan, B., and Schiffer, C. (1993). A randomized clinical trial comparing Melphalan/Prednisone with and without $\alpha - 2b$ interferon in newly-diagnosed patients with multiple myeloma: A cancer and leukemia group B study. *Journal of Clinical Oncology* **11**, 155-160.

18. Cooper, H., Hedges, L.V. (1994). *The Handbook Research Synthesis.* Russell Sage Foundation, New York.

19. Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1-38.

20. DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.

21. DuMouchel, W. and Normand, S.-L. (2000). Computer-modeling and graphical strategies for meta-analysis. In: *Meta-Analysis in Medicine and Health Policy*, ed. by D. K. Stangl and D. A. Berry, Marcel Dekker, Basel, 127-178.

22. Goldstein, H. (1995). *Multilevel Statistical Models.* Arnold, London.

23. Greenland, S., Robins, J.M. (1985). Estimation of common effect parameter from sparse follow up data. *Biometrics* **41**, 55-68.

24. Hedges, L.V. (1994). Combining estimates across studies: Meta-analysis of research. In: K.A. McCormick, S.R. Moore, R.A. Siegel (Eds.) Clinical Practice Guideline Development: Methodology Perspectives. Washington, D. C.: Agency for Health Care Policy Research, 15-26.

25. Hine, L.K., Laird, N., Hewitt, P. and Chalmers, T.C. (1989). Meta-analytic evidence against prophylactic use of lidocaine in myocardial infarction. *Archives of Internal Medicine* **149**, 2694-2698.

26. Horvitz, R. I., Singer, B. H., Makuch, R. W., Viscoli, C. M. (1996). Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. *Journal of Clinical Epidemiology* **49**, 395-400.

27. Laird N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805-811.

28. Lange, K. (2004). *Optimization.* Springer, New York, Berlin, London.

29. Le, C.T. (1992). *Fundamentals of Biostatistical Inference.* Marcel Dekker, Inc., New York, Basel, Hong Kong.

30. Lindsay, B.G. (1983). The geometry of mixture likelihoods, Part I: A general theory. *Annals of Statistics*, **11**, 783–792.

31. Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry, and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Statistical Mathematics, Hayward.

32. McLachlan, G. and Krishnan, T. (1997). *The EM Algorithm and Extensions.* Wiley, New York.

33. McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman & Hall, London, 2nd Edition.

34. McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.

35. Murphy, S.A. and Van der Vaart, A.W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95,** 449–85.

36. Patefield, W.M., (1977). On the maximized likelihood function. *Sankhya, Series B* **39**, 92-96.

37. Pocock, S.J. (1997). *Clinical Trials: A Practical Approach.* John Wiley & Sons, Chichester, New York, Brisbane

38. Thompson, S.G., Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* **18**, 2693-2708.

39. Woodward, M. (1999). *Epidemiology: Study Design and Data Analysis.* Chapman & Hall/CRC, Londeon, New York, Washigton, D.C.

40. Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases* **27**, 335-371.

Table 1: Data Illustration for a Multicenter Clinical Trial for Studying the Effect of Beta-Blocker for Reducing Mortality after Myocardial Infarction (Yusuf *et al.* 1985)

Center i	Deaths x_i^T	Person-Time n_i^T	Deaths x_i^C	Person-Time n_i^C
1	3	38	3	39
2	7	114	14	116
3	5	69	11	93
4	102	1533	127	1520
5	28	355	27	365
6	4	59	6	52
7	98	945	152	939
8	60	632	48	471
9	25	278	37	282
10	138	1916	188	1921
11	64	873	52	583
12	45	263	47	266
13	9	291	16	293
14	57	858	45	883
15	25	154	31	147
16	33	207	38	213
17	28	251	12	122
18	8	151	6	154
19	6	174	3	134
20	32	209	40	218
21	27	391	43	364
22	22	680	39	674

Table 2: Outcome Data for Prophylactic Use of Lidocaine after Heart Attack (AMI) (Hine *et al.* 1989, following Normand 1999)

Center i	Deaths x_i^T	n_i^T	Deaths x_i^C	n_i^C	$\hat{\theta}_i$(95%CI)
1	2	39	1	43	2.21 (0.21-23.4)
2	4	44	4	44	1.00 (0.27-3.75)
3	6	107	4	110	1.54 (0.45-5.31)
4	7	103	5	100	1.36 (0.45-4.14)
5	7	110	3	106	2.25 (0.60-8.47)
6	11	154	4	146	2.61 (0.85-8.01)

Table 3: Outcome Data of a Meta-Analysis of Smith *et al.* (1993) on the Effect of Cholesterol Lowering Treatment on Mortality from Coronary Heart Disease (following Arends *et al.* 2000)

Center i	Deaths x_i^T	n_i^T	Deaths x_i^C	n_i^C
1	28	380	51	350
2	70	1250	38	640
3	37	690	40	500
4	2	90	3	30
5	0	30	3	30
6	61	1240	82	1180
7	41	1930	55	890
8	20	340	24	350
9	111	1930	113	1920
10	81	1240	27	410
11	31	1140	51	1140
12	17	210	12	220
13	23	210	20	230
14	0	90	4	170
15	1450	38620	723	19420
16	174	1350	178	1330
17	28	890	31	860
18	42	1970	48	2060
19	4	150	5	150
20	37	2150	48	2100
21	39	1010	28	1120
22	8	100	1	50
23	5	340	7	340
24	269	4410	248	4390
25	49	3850	62	3740
26	0	190	1	190
27	19	1510	12	1560
28	68	13850	71	13800
29	46	10140	43	10040
30	33	5910	3	1500
31	236	27630	181	27590
32	0	100	1	100
33	1	20	2	30

Table 4: Outcome Data For Treatment Group of a Multicenter Clinical Trial (With High Sparsity) (Cancer and Leukaemia Group, Cooper *et al.* 1993)

Center i	Deaths x_i^T	n_i^T	Deaths x_i^C	n_i^C
1	1	3	3	4
2	8	11	3	4
3	2	3	2	2
4	2	2	2	2
5	0	3	2	2
6	2	3	1	3
7	2	3	2	2
8	4	4	1	5
9	2	3	2	2
10	2	3	0	2
11	3	3	3	3
12	0	2	2	2
13	1	5	1	4
14	2	4	2	3
15	4	6	2	4
16	3	9	4	12
17	2	3	1	2
18	1	4	3	3
19	2	3	1	4
20	0	2	0	3
21	1	5	2	4

Table 5: Results of the mixture model fitting for the multicenter trial of Cholesterol Lowering Treatment and Coronary Heart Disease given in Table 3; $\hat{Q}^{(m)}$ is the mixture maximum profile likelihood estimate with m components

m	$\hat{\theta}_j$	$\hat{q}_j$	$d(\theta_{max}, \hat{Q}^{(m)})$	$L^*(\hat{Q}^{(m)})$	BIC
1	0.9716	1	10,518.11	-50,172.01	-100,347.52
2	1.0058	0.8901	1.7856	-50,161.54	-100,333.57
	0.4401	0.1099			
3	0.9776	0.8029	1.2998	-50,160.62	-100,338.72
	0.4283	0.1013			
	1.2827	0.0958			
4	1.0016	0.6546	1.0000	-50.159,66	-100,343.79
PNMLE	0.3665	0.0558			
	0.6962	0.1916			
	1.2793	0.0980			

Table 6: Classification of centers for the multicenter Cholesterol Lowering Treatment trial given in Table 3 according to the posterior distribution when using the two-component mixture $\hat{Q}^{(2)}$ (see Table 5)

Center	posterior of component 1	2	classified as belonging to
1	0.09775	0.902248	2
2	0.99993	0.000069	1
3	0.8994	0.100599	1
4	0.73118	0.268823	1
5	0.78973	0.21027	1
6	0.97666	0.023345	1
7	0.00002	0.999979	2
8	0.98652	0.01348	1
9	1	0	1
10	0.99993	0.000068	1
11	0.63296	0.367043	1
12	0.99889	0.001115	1
13	0.99951	0.000493	1
14	0.82041	0.179595	1
15	1	0	1
16	1	0	1
17	0.99498	0.005023	1
18	0.99958	0.000422	1
19	0.91806	0.081936	1
20	0.98398	0.016024	1
21	1	0.000001	1
22	0.98862	0.011379	1
23	0.90455	0.09545	1
24	1	0	1
25	0.99441	0.005587	1
26	0.88153	0.118473	1
27	0.99956	0.000443	1
28	1	0.000005	1
29	0.99997	0.000029	1
30	0.99963	0.000369	1
31	1	0	1
32	0.88138	0.118618	1
33	0.89578	0.104223	1

Table 7: Results of fitting various models to the multicenter trial of Smoking Cessation

$L^*(\hat{\beta})$	Covariates	$\hat{\beta}_j$	S.E.	P-Value
-17,218.81	Intercept	0.4483	0.0392	0.0000
-17,218.73[§]	Intercept	0.4700	0.0647	0.0000
	High Support	-0.0343	0.0813	0.3367
-17,215.49[§]	Intercept	0.3850	0.0459	0.0000
	Patch	0.2301	0.0887	0.0047
-17,214.84	Intercept	0.4356	0.0661	0.0000
	Patch	0.2526	0.0912	0.0028
	High Support	-0.0803	0.0838	0.1434
-17,214.65	Intercept	0.4222	0.0697	0.0000
	Patch	0.3558	0.1950	0.0340
	High Support	-0.0657	0.0926	0.2391
	Patch×H.S.	-0.1328	0.2207	0.2738

[§]Note that these models are not nested.

Table 8: NPMLE $\hat{Q}^{(15)}$ for the Multi-Level Modelin the MAIPD of Cholesterol Lowering Treatment and Coronary Heart Disease given in Table 3; H is the Number of Studies Allocated to the Associated Component

	Component-Number in Multi-Level-Model					
Component	1.	2.	3.	4.	5.	6.
baseline α	-3.292866	-5.355982	-2.832449	-2.016804	-3.767197	-5.022077
risk ratio θ	1.007591	1.022630	1.020841	.965525	.817292	1.306739
weight q	.077422	.104689	.226971	.034156	.113021	.056061
H	2	4	10	1	5	1
Component	7.	8.	9.	10.	11.	12.
baseline α	-2.789155	-1.931629	-4.086842	-2.458914	-3.661684	-4.850073
risk ratio θ	.347178	.506721	.759630	1.258181	1.481572	1.587610
weight q	.045034	.033068	.058496	.028397	.035870	.023181
H	1	1	1	1	1	1
Component	13.	14.	15.			
baseline α	-3.152536	-2.664810	-6.166033			
risk ratio θ	.654683	.766163	2.652461			
weight q	.055444	.087982	.020206			
H	1	2	1			

$L^* = -238.911862$
$\max_{\alpha,\beta} d(\alpha, \beta, Q) = 1.000000$

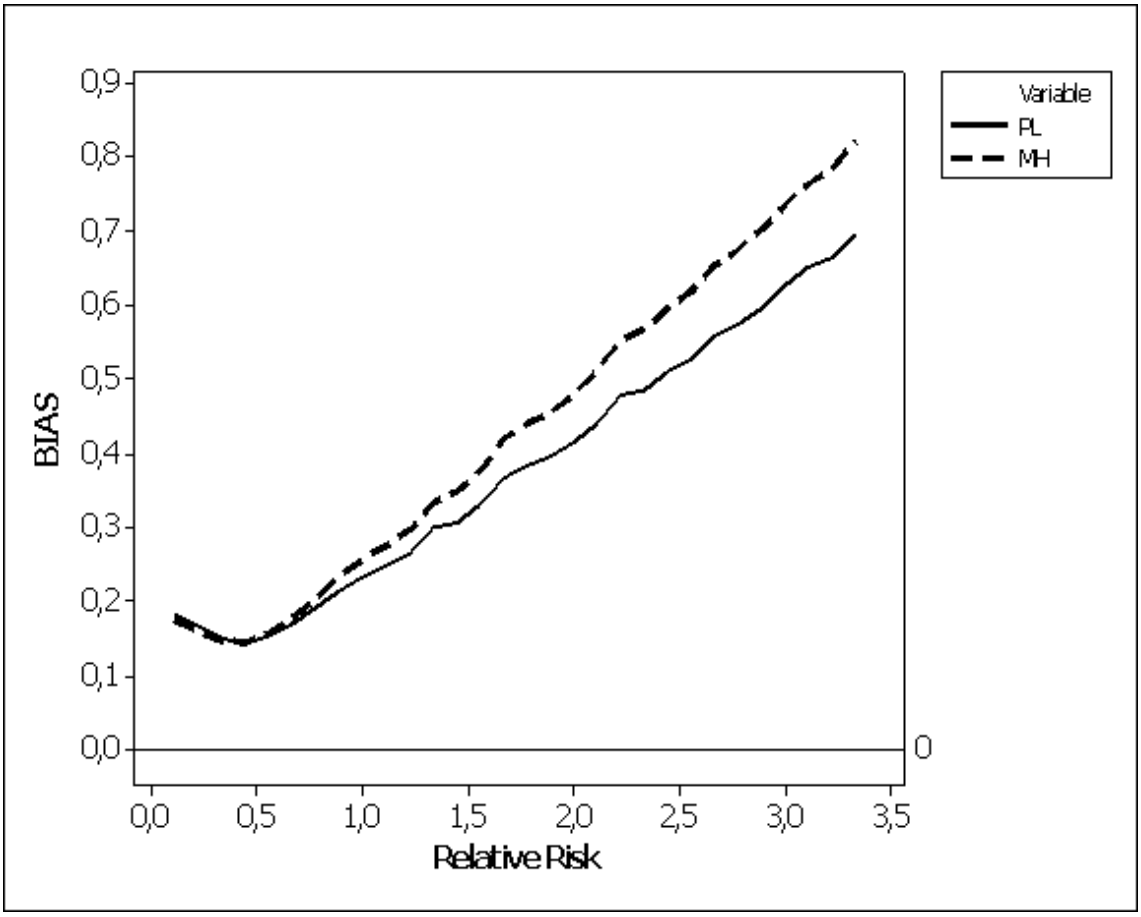

Figure 1: *A Comparison of Profile Maximum Likelihood Estimator (PL) and Mantel-Haenszel Estimator (MH) for Sparse Multicenter Trial with Respect to Bias based upon a Simulation*

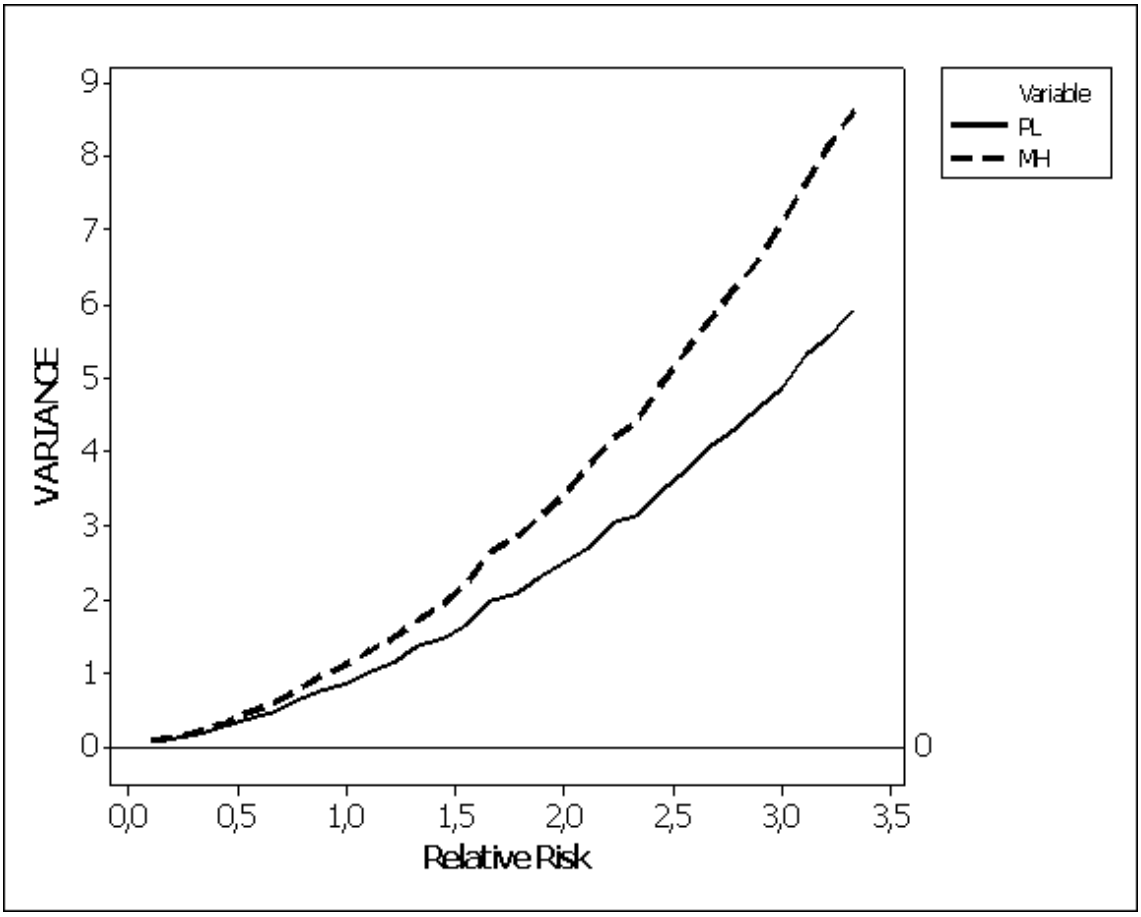

Figure 2: *A Comparison of Profile Maximum Likelihood Estimator (PL) and Mantel-Haenszel Estimator (MH) for Sparse Multicenter Trial with Respect to Variance based upon a Simulation*

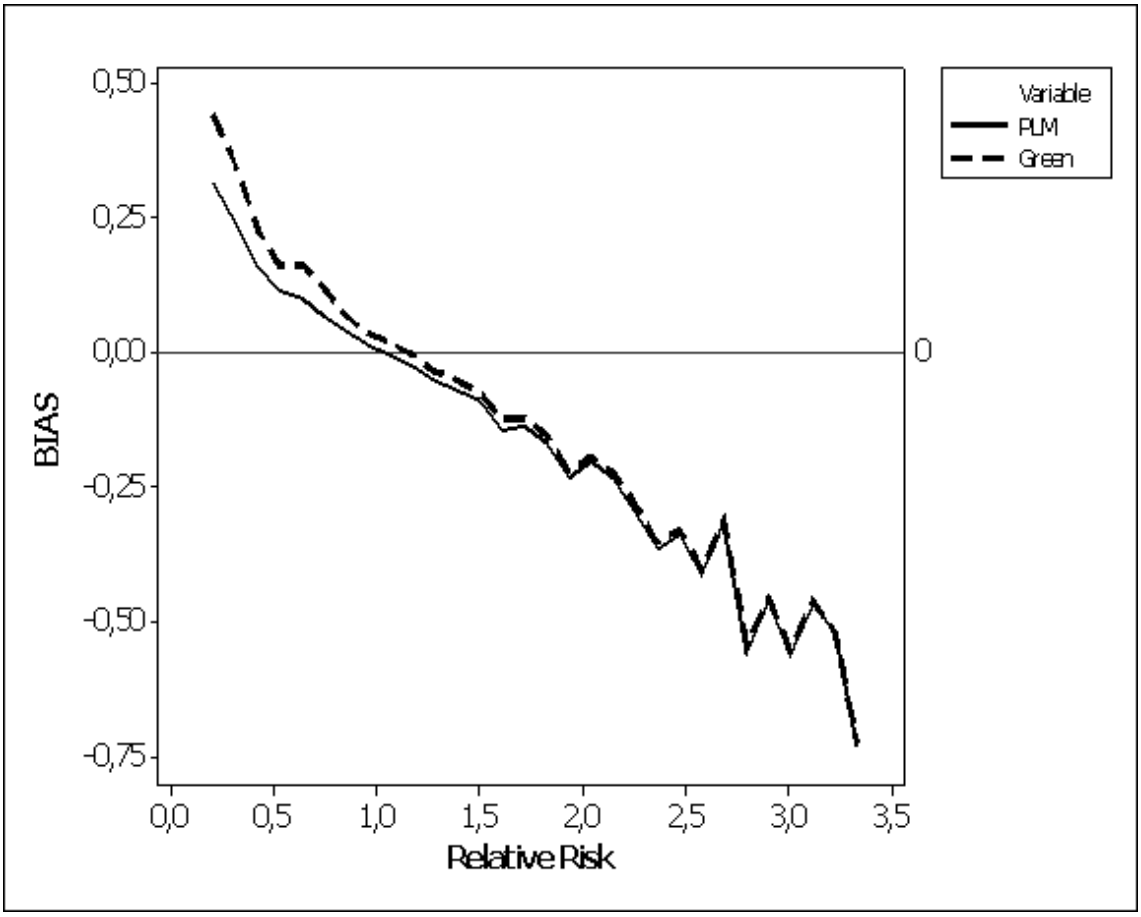

Figure 3: *A Comparison of Variance Formulas provided by (19) (PML) and Greenland and Robbins (20) (Green) for Sparse Multicenter Trial with Respect to Bias based upon a Simulation*

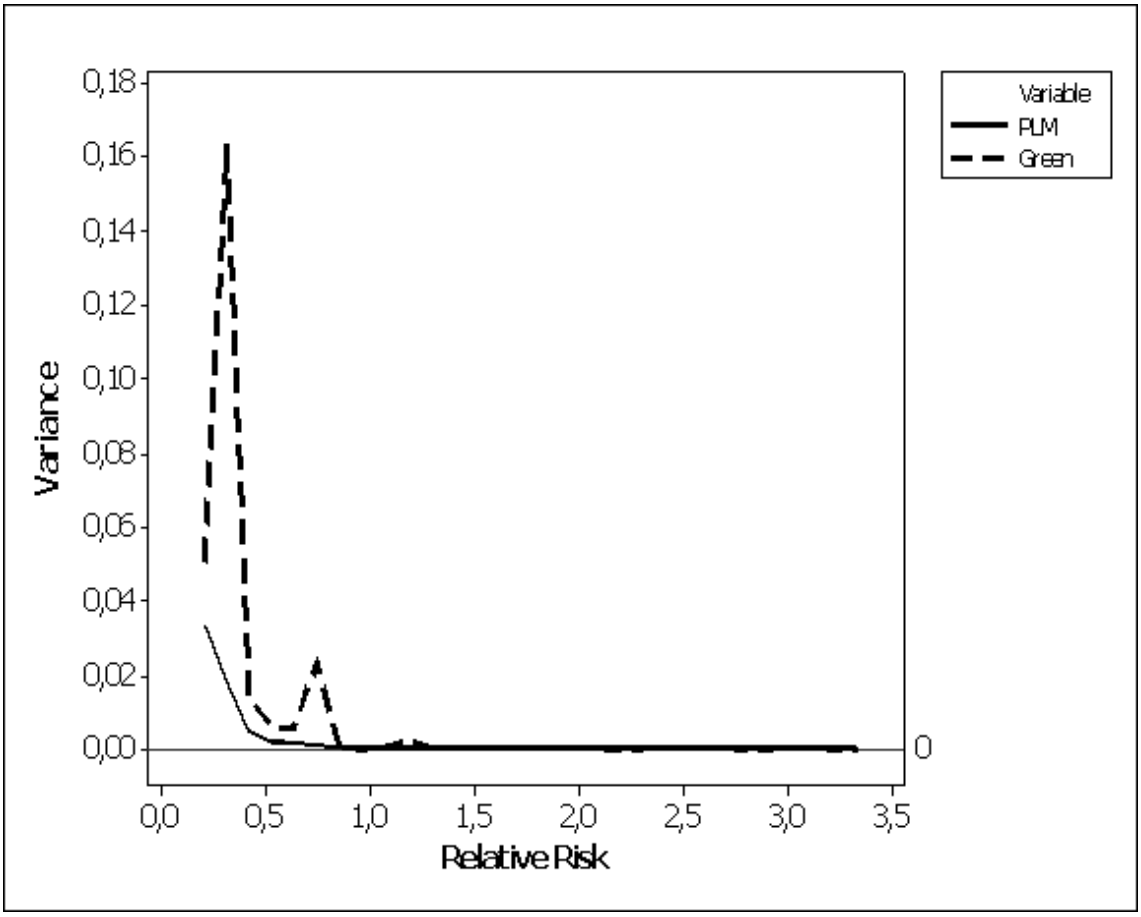

Figure 4: *A Comparison of Variance Formulas provided by (19) (PML) and Greenland and Robbins (20) (Green) for Sparse Multicenter Trial with Respect to Variance based upon a Simulation*

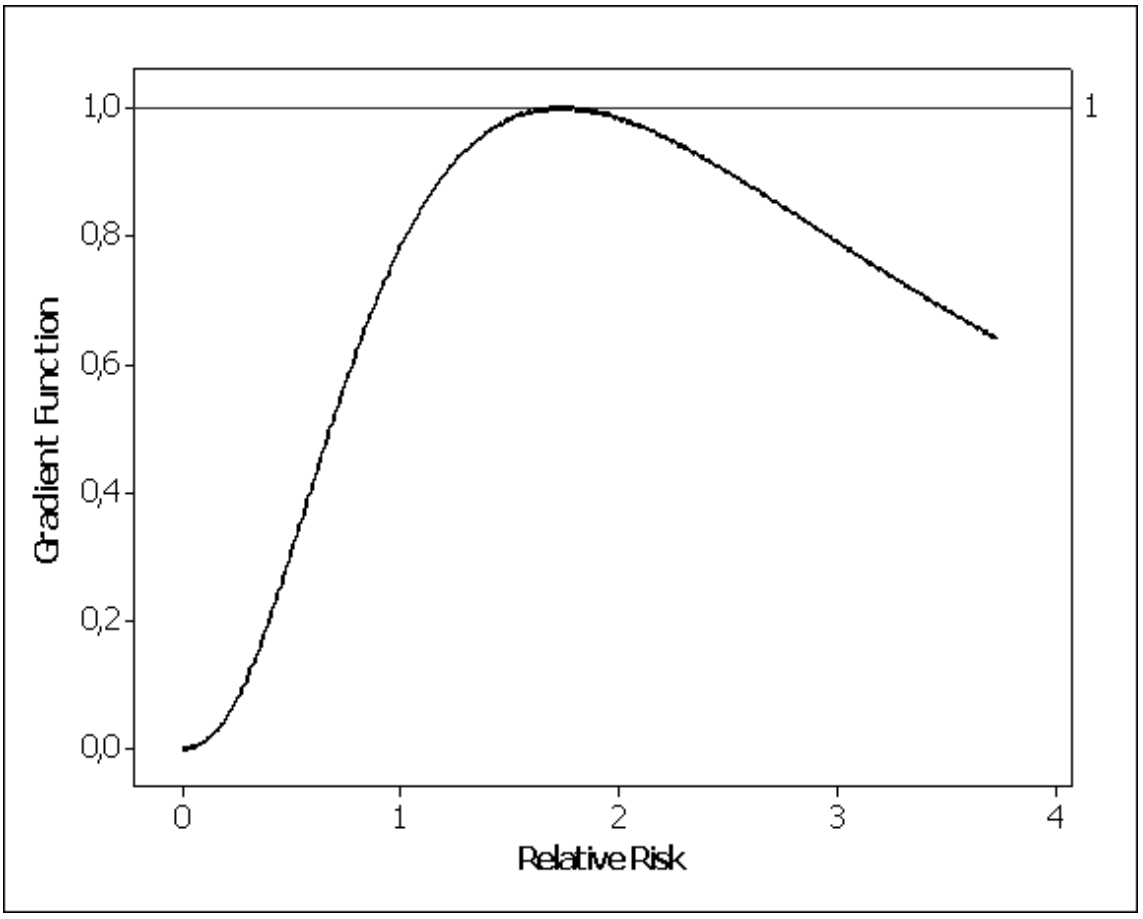

Figure 5: *Gradient function $d(\theta, \hat{\theta}_{PMLE})$ for Lidocaine trial (see also Table 2)*

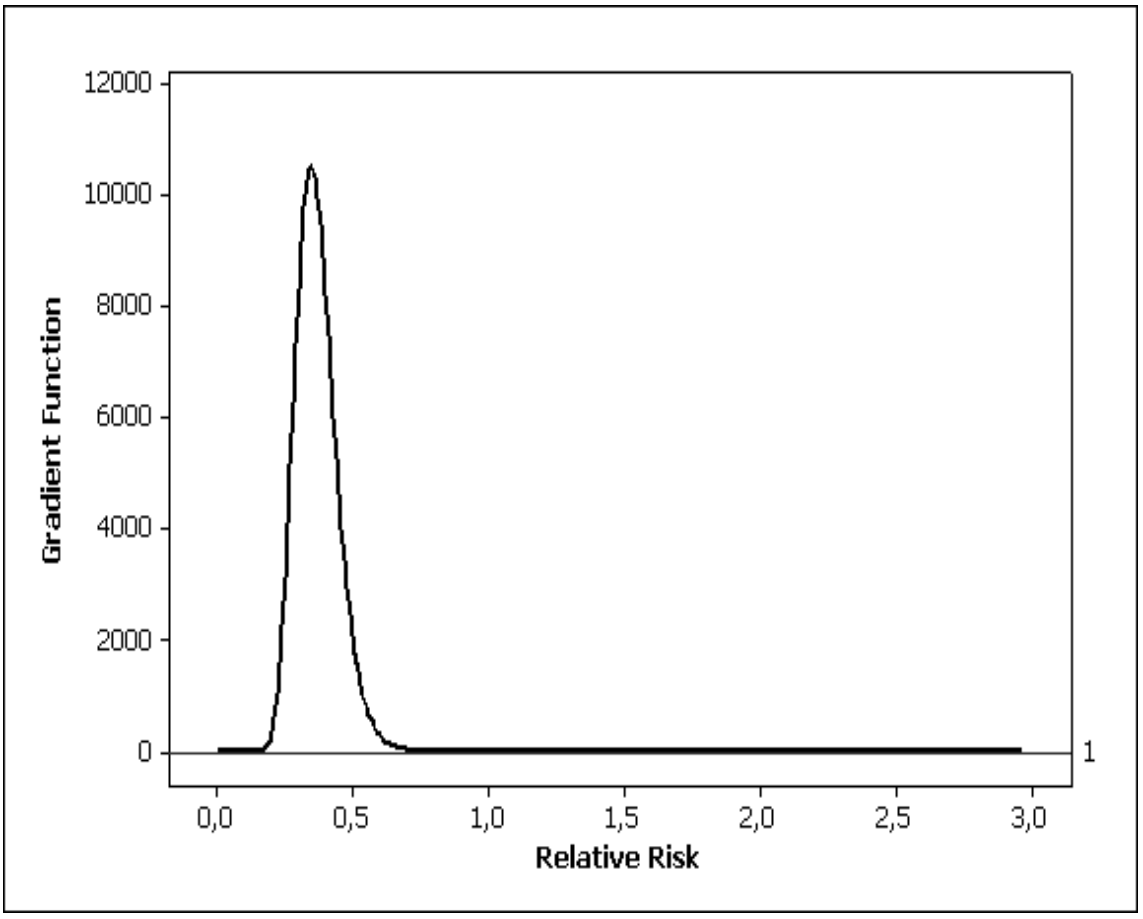

Figure 6: *Gradient function $d(\theta, \hat{\theta}_{PMLE})$ for Cholesterol Lowering trial (see also Table 3)*

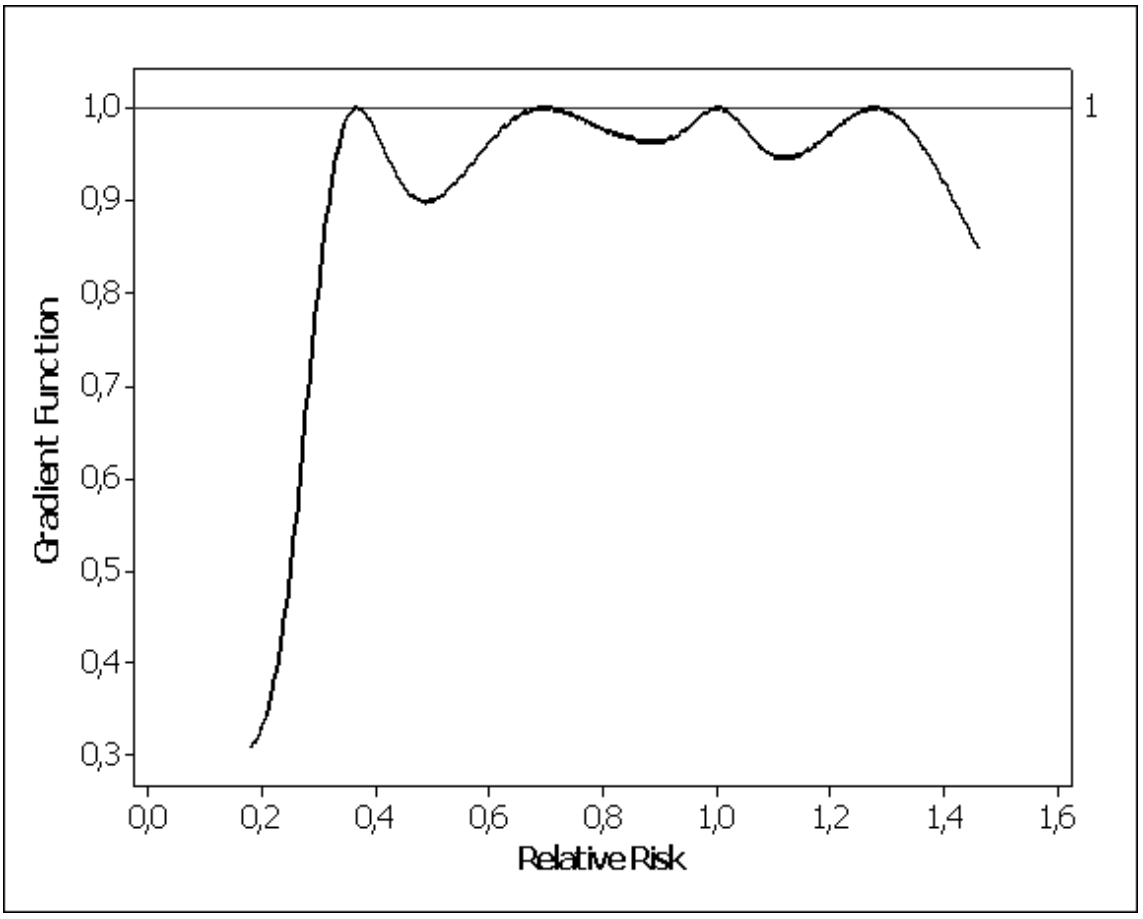

Figure 7: *Gradient function $d(\theta, \hat{Q}_{PNMLE})$ (see also Table 5) for Cholesterol Lowering trial (see also Table 3)*

Table 1 Fifty-nine Nicotine Replacement Therapy (NRT) Trials Examining the Efficacy of NRT on Smoking Cessation*

ID	Study Name	Year	Treated y_T/n_T	Control y_C/n_C
	Gum: high-intensity support			
1	Puska	1979	29/116	21/113
2	Malcolm	1980	6/73	3/121
3	Fagerstrom	1982	30/50	23/50
4	Fee	1982	23/180	15/172
5	Jarvis	1982	22/58	9/58
6	Hjalmarson	1984	31/106	16/100
7	Killen	1984	16/44	6/20
8	Schneider	1985	9/30	6/30
9	Hall	1987	30/71	14/68
10	Tonnesen	1988	23/60	12/53
11	Blondal	1989	37/92	24/90
12	Garcia	1989	21/68	5/38
13	Killen	1990	129/600	112/617
14	Nakamura	1990	13/30	5/30
15	Campbell	1991	21/107	21/105
16	Jensen	1991	90/211	28/82
17	McGovern	1992	51/146	40/127
18	Pirie	1992	75/206	50/211
19	Zelman	1992	23/58	18/58
20	Herrera-1	1995	37/76	17/78
	Patch: high-intensity support (contd.)			
31	Richmond	1994	40/160	19/157
32	Kornitzer	1995	19/150	10/75
33	Stapleton	1995	77/800	19/400
34	Campbell	1996	24/115	17/119
	Gum: low-intensity support			
35	BR SOCIETY	1983	39/410	111/1208
36	Russell	1983	81/729	78/1377
37	Fagerstrom	1984	28/106	5/49
38	Jamrozik	1984	10/101	8/99
39	Jarvik	1984	7/25	4/23
40	Clavel-Chapel	1985	24/205	6/222
41	Schneidera	1985	2/13	2/23
42	Page	1986	9/93	13/182
43	Campbell	1987	13/424	9/412
44	Sutton	1987	21/270	1/64
45	Areechon	1988	56/99	37/101
46	Harackiewicz	1988	12/99	7/52
47	Llivina	1988	61/113	28/103
48	Sutton	1988	5/79	2/82
49	Gilbert	1989	11/112	9/111

Table 1 (Cont'd)

ID	Study Name	Year	Treated y_T/n_T	Control y_C/n_C
	Patch: high-intensity support			
21	Buchkremer	1981	11/42	16/89
22	Hurt	1990	8/31	6/31
23	Ehrsam	1991	7/56	2/56
24	Tnsg	1991	111/537	31/271
25	Sachs	1993	28/113	10/107
26	Westman	1993	16/78	2/80
27	Fiore-1	1994	15/44	9/43
28	Fiore-2	1994	10/57	4/55
29	Hurt	1994	33/120	17/120
30	ICRF	1994	76/842	53/844
	Gum: low-intensity support (contd.)			
50	Hughes	1989	23/210	6/105
51	Hughes	1990	15/59	5/19
52	Mori	1992	30/178	22/186
53	Nebot	1992	5/106	13/319
54	Fortmann	1995	44/262	42/261
	Patch trials: low-intensity support			
55	Abelin	1989	17/100	11/99
56	Daughton	1991	28/106	4/52
57	Tonneson	1991	17/145	2/144
58	Burton	1992	29/115	22/119
59	Paoletti	1996	15/60	4/60

*The numbers quitting and enrolled in the treatment (T) and control (C) groups are denoted y_T, n_T, y_C, and n_C, respectively.

Figure 8: *An Example of a Multicenter Study with Three Covariates (from DuMouchel and Normand 2000)*

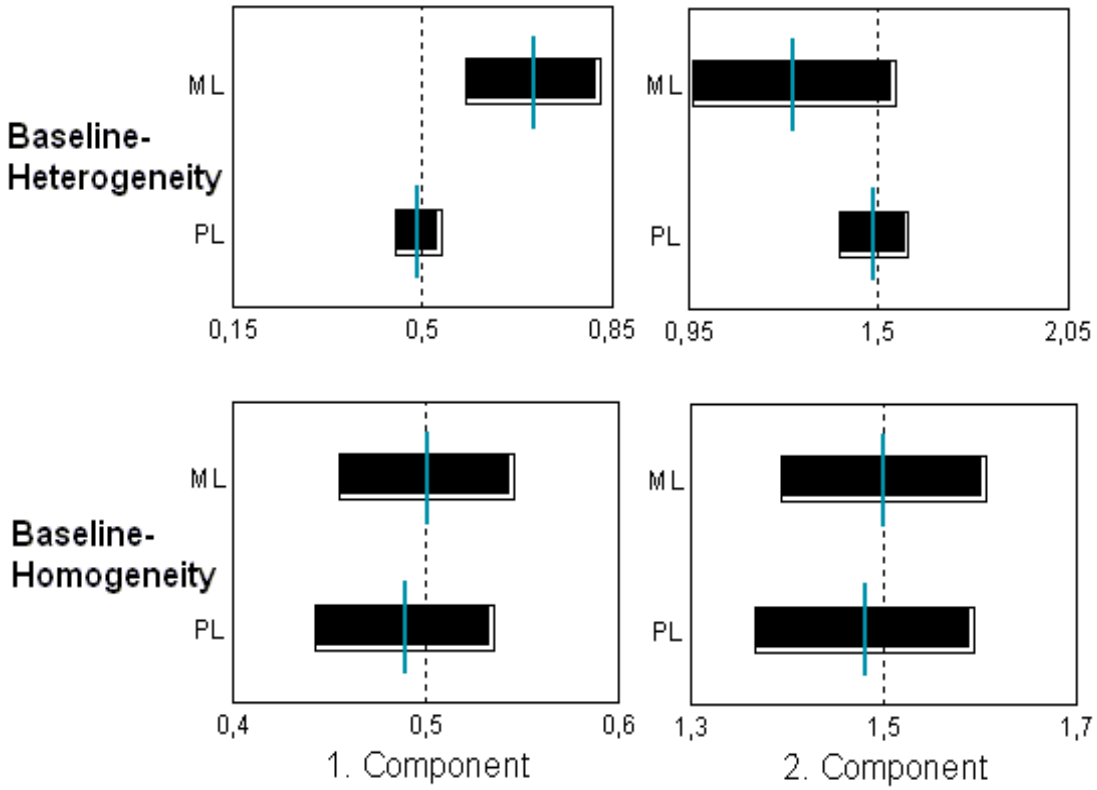

Figure 9: *Comparing the Profile Likelihood Approach (PL) with the Multi-Level Model by Means of a Simulation Study with Two-Component Treatment Heterogeneity and under Baseline Heterogeneity (upper panel) and Homogeneity (bottom panel)*

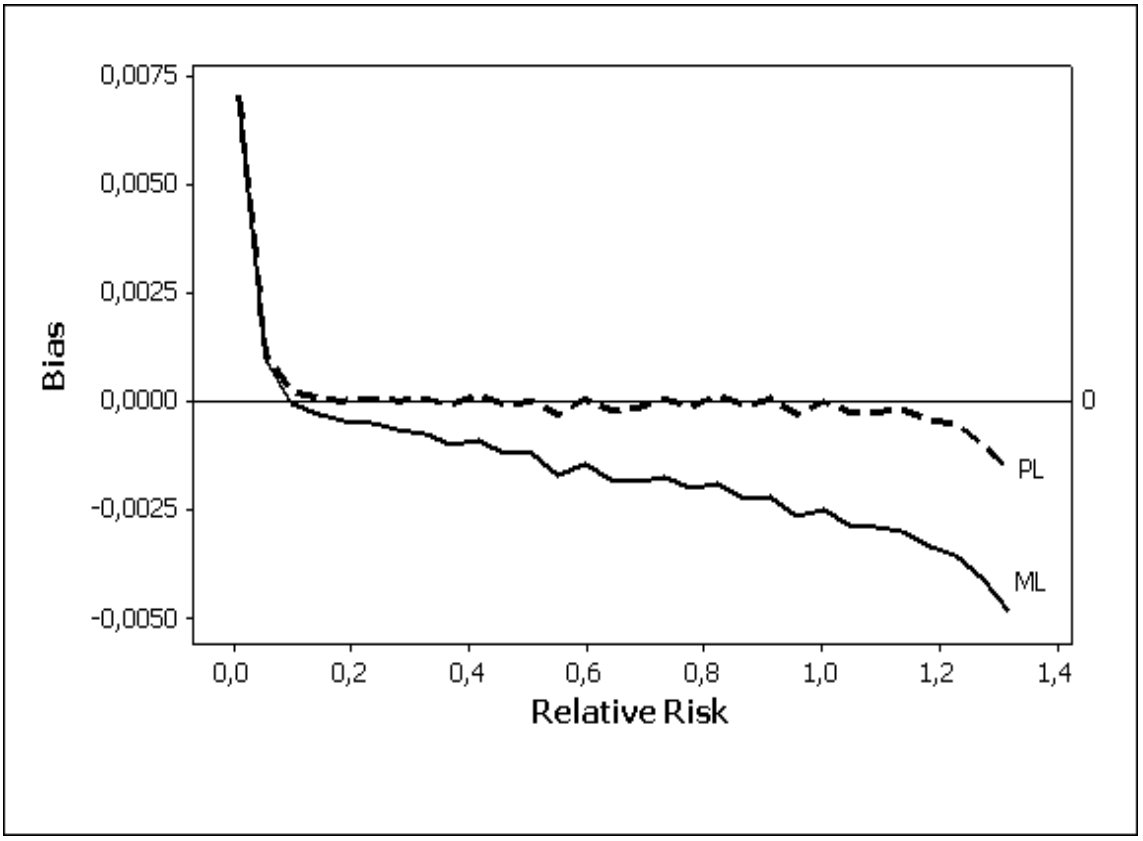

Figure 10: *Comparing the Bias of the Profile Likelihood Approach (PL) with the Multi-Level Model by Means of a Simulation Study with Treatment Homogeneity and Baseline Heterogeneity*