Approximate Conditional Inference in Mixed Effects Models with Binary Data

Woojoo Lee
Department of Statistics
Seoul National University, Korea

Jian Qing Shi*
School of Mathematics & statistics
University of Newcastle, UK

Youngjo Lee
Department of Statistics
Seoul National University, Korea

August 24, 2007

*Address for correspondence: School of Mathematics & statistics, University of Newcastle, NE1 7RU, UK. Email: j.q.shi@ncl.ac.uk

Summary

Conditional likelihood approach is a sensible choice for a hierarchical logistic regression model or other generalized regression models with binary data. However, its heavy computational burden limits its use, especially for the related mixed effects model. In this paper, we use modified profile likelihood as an accurate approximation to conditional likelihood, and then propose the use of two methods for inferences for the hierarchical generalized regression models with mixed effects. One is based on hierarchical likelihood and Laplace approximation method, and the other is based on Markov chain Monte Carlo EM algorithm. The methods are applied to a meta-analysis model for trend estimation and the model for multi-arm trials. A simulation study is conducted to illustrate the performance of the proposed methods.

Keywords: Conditional likelihood, Hierarchical likelihood, Laplace approximation, Meta-analysis model, MCMC-EM algorithm, Mixed effects model, Modified profile likelihood, Multi-arm trails.

1 Introduction

We consider a meta-analysis model combining several 2×2 tables. Suppose that a typical 2×2 table has m_t cases and m_c controls, and binomial outcomes

$$Z_{ti} \sim Bin(m_{ti}, \pi_{ti}); \quad Z_{ci} \sim Bin(m_{ci}, \pi_{ci}); \quad i = 1, \ldots, K.$$

The log odds ratio for this study is

$$\eta_i = \log\left(\frac{\pi_{ti}}{1-\pi_{ti}} \cdot \frac{1-\pi_{ci}}{\pi_{ci}}\right),$$

which is the parameter of a main interest. When m_{ti} and m_{ci} are sufficiently large and when π_{ti} and π_{ci} are not very close to zero or one, we can use normal model based on empirical logistic transformations (Cox and Snell, 1989); otherwise, we should use the above exact binomial distributions. Unconditional approach is based on the above binomial distribution with

$$\log\left(\frac{\pi_{ci}}{1-\pi_{ci}}\right) = \alpha_i, \quad \log\left(\frac{\pi_{ti}}{1-\pi_{ti}}\right) = \alpha_i + \eta_i,$$

where α_i stands for trial effects. The model with $\eta_i = \eta$ is called fixed effect model. When m_{ti} and m_{ci} are small and K is large, the estimation based on unconditional likelihood approach may be biased (Lubin, 1981 and Cox and Snell, 1989). For example, the unconditional maximum likelihood estimator $\hat{\eta}$ is close to 2η (Cox and Snell, 1989, p59 and p103) for matched pairs, having $m_{ti} = m_{ci} = 1$. The conditional likelihood approach gives a better estimator, but involving heavy computation (see discussion in Liao, 1999 and Vollset, Hirji and Elashoff, 1991). Therefore, a development of an efficient computational method for conditional approach is essential. Among many others, Davison (1988) proposed a method based on saddlepoint approximation (see also Barndorff-Nielsen and

Cox, 1979); Sartori (2003) used modified profile likelihoods as an accurate approximations to conditional likelihood; and Shi and Copas (2002) proposed a method based on Markov chain Monte Carlo EM (MCMC-EM) algorithm.

This paper is concerned with the conditional inference in mixed effects models with binary data. The difficulty here is that the conditional likelihood involves an intractable integration which the integrand is the conditional density function for binomial distribution. The difficulty increases as the dimensionality of integration increases as for example the meta-analysis models for multi-arm trials we consider in Section 4. Our basic idea is to approximate the conditional density by modified profile likelihood and then use Laplace approximation to the marginal likelihood. The other method is to use a MCMC-EM algorithm. Section 2 describes conditional inference for a logistic regression model with random effects and its accurate approximation. Sections 3 and 4 apply the proposed methods to meta-analysis model for trend estimation and the model for multi-arm trials. Some simulation study results are reported in Section 5, and some discussion is given in Section 6.

2 Mixed Effects Models with Binary Data

2.1 Conditional Inference

We now consider a hierarchical logistic regression model. Suppose that there are several sets of data, in which the ith data set has $(n_i + 1)$ treatment groups, and has z_{ij} cases out of m_{ij} subjects in jth treatment group for $j = 0, 1, \ldots, n_i$, where $j = 0$ stands for the

base-line control group. The binomial outcomes have the distribution:

$$z_{ij} \sim Bin(m_{ij}, \pi_{ij}), \quad j = 0, 1, \ldots, n_i. \tag{1}$$

The related log-odds is defined as

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \tag{2}$$

for $j = 0, 1, \cdots, n_i$. The ith set of data can be the data collected in ith study in meta-analysis, the data collected in the ith centre in multi-centre analysis or the data collected in ith subject in trend estimation. If the jth treatment group is associated with a univariate covariate x_{ij} (e.g. the dose level in trend estimaton) for ith set of data (usually $x_{i0} = 0$), we can define a logistic regression model by

$$\eta_{ij} = \alpha_i + \beta_i x_{ij}, \quad j = 0, 1, \ldots, n_i \tag{3}$$

where β_i measures the association between the log-odds ratio and x_{ij}, the effect of covariate such as dose level in trend estimation which is the parameter of interest, and α_i is a nuisance parameter.

When m_{ij} is relatively large, we can use a normal approximation for empirical log-odds ratio (Shi and Copas, 2004). However, such a normal approximation may not be appropriate when some m_{ij} is small, while K is large, so that we may use an exact likelihood. There are problems to use an unconditional approach for the models. If α_i is fixed, we have not got enough data in each trial to estimate α_i; we face the notorious problem of *infinitely many nuisance parameters* that leads to inconsistent estimates (Andersen, 1970). An alternative way is to assume a random-effect model for α_i's:

$$\alpha_i \sim H(\cdot, \theta_\alpha).$$

However, it is often not feasible to find such a model, for example, when the data includes both retrospective and prospective trials in meta analysis. Even if we can assume a distribution, the inference on parameters of interest could be sensitive to the selection of the distributional assumption about α_i.

Following the above discussion and the discussion given in Section 1, a sensible way is to use a conditional approach to eliminate nuisance parameter α_i. A direct way is to consider the conditional likelihood given $T_i = Z_{i0}+Z_{i1}+\cdots+Z_{in_i} = t_i = z_{i0}+z_{i1}+\cdots+z_{in_i}$:

$$\begin{aligned} & f(\boldsymbol{z}_i|T_i = t_i;\beta_i) = p(\boldsymbol{Z}_i = \boldsymbol{z}_i|Z_{i0} + Z_{i1} + \cdots + Z_{in_i} = t_i) \\ = & \frac{\prod_{j=0}^{n_i} \left(\begin{array}{c} m_{ij} \\ z_{ij} \end{array} \right) \exp(\beta_i z_{ij} x_{ij})}{\sum_{u_i \in \mathcal{U}_i} \left(\begin{array}{c} m_{i0} \\ t_i - u_{i1} - \cdots - u_{in_i} \end{array} \right) \prod_{j=1}^{n_i} \left(\begin{array}{c} m_{ij} \\ u_{ij} \end{array} \right) \exp(\beta_i u_{ij} x_{ij})}, \end{aligned} \tag{4}$$

where $u_i = (u_{i1}, \cdots, u_{in_i})^t$, and

$$\mathcal{U}_i = \{u_i : 0 \le u_{ij} \le m_{ij}, \text{ and } t_i - m_{i0} \le u_{i1} + \cdots + u_{in_i} \le t_i\}.$$

By the conditional approach the nuisance parameters α_i are eliminated. However, the denominator, which is proportional to $\mathrm{P}(T_i = t_i)$, is hard to enumerate unless n_i is small. Furthermore, the combination term $\left(\begin{array}{c} m_{ij} \\ u_{ij} \end{array} \right)$ cannot be computed with ordinary statistical packages when m_{ij} is moderately large (see for example Vollset, Hirji and Elashoff, 1991). Thus, in this paper we first study how to approximate (4) accurately to implement conditional inferences.

2.2 Modified profile likelihood

Barndorff-Nielsen and Cox (1994, p288) proposed using the modified profile likelihood to approximate the conditional likelihood for 2×2 tables. It reduces the biases of estimators

from profile likelihood in general.

Consider the parametric model for data $\boldsymbol{z}$, with parameter θ=(β, α) and log-likelihood

$$\ell(\beta, \alpha) = \ell(\beta, \alpha; \boldsymbol{z}) = \log f(\boldsymbol{z}; \beta, \alpha).$$

When the data $\boldsymbol{z}$ can be replaced in the likelihood by $(\hat{\beta}, \hat{\alpha}, A)$, where A is ancillary statistic, so that

$$\ell(\beta, \alpha; \hat{\beta}, \hat{\alpha}, A)$$

is proportional to $\ell(\beta, \alpha; \boldsymbol{z})$.

First consider the profile likelihood,

$$\ell^P(\beta) = \ell(\beta, \hat{\alpha}_\beta),$$

where $\hat{\alpha}_\beta$ is the maximum likelihood (ML) estimator of α when the value of β is treated as fixed. However, this profile likelihood could give biased estimation, so that the modified profile likelihood below can be used

$$\ell^M(\beta) = \ell^P(\beta) + M(\beta)$$

where

$$M(\beta) = -\log|\ell_{\alpha;\hat{\alpha}}(\beta, \hat{\alpha}_\beta; \hat{\beta}, \hat{\alpha}, A)| + \frac{1}{2}\log|j_{\alpha\alpha}(\beta, \hat{\alpha}_\beta; \hat{\beta}, \hat{\alpha}, A)|,$$

$|\cdot|$ is determinant,

$$\ell_{\alpha;\hat{\alpha}} = \partial^2\ell(\beta, \alpha; \hat{\beta}, \hat{\alpha}, A)/\partial\alpha\partial\hat{\alpha}^t$$

is the matrix of mixed second order derivatives and $j_{\alpha\alpha} = -\partial^2\ell(\beta, \alpha; \hat{\beta}, \hat{\alpha}, A)/\partial\alpha\partial\alpha^t$ is the observed information matrix with respect to α (Barndorff-Nielsen and Cox (1994)). Under appropriate regularity conditions, the modified profile likelihood $\ell^M(\beta)$ approximates

both conditional and marginal likelihood, when they exist (Severini, 2000, Section 9.3.2 and 9.3.3). Furthermore, the modified profile likelihood is quite effective even when neither a conditional nor marginal likelihood exists, since the resulting estimating equation is approximately unbiased (Pace and Salvan, 1997, Section 11.6). The modified profile likelihoods is invariant with respect to reparametrizations of α, while the profile likelihood is not.

Sartori (2003) gives a theoretical comparison of the asymptotic properties of profile and modified profile likelihoods in stratified model, considering a setting in which both the number of strata, K, and the average stratum sample size, $\bar{n} = n/K$, increase to infinity. In particular, a sufficient condition for the normal approximation to the distribution of usual likelihood based statistics is $K = o(\bar{n}) = o(n^{1/2})$ for the profile likelihood, and $K = o(\bar{n}^3) = o(n^{3/4})$ for the modified profile likelihood. Moreover, even when the condition does not hold, the modified profile likelihood gives a remarkable improvement over profile likelihood in terms of bias of estimates. The simulation study in Sartori (2003) showed that the approximation of conditional likelihood by using modified profile likelihood is accurate even if m is very small comparing to K for binary data.

Following Severini (2000) we can show (the proof is given in Appendix) that in the logistic regression models (1) to (3), i.e.,

$$\eta_{ij} = \log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \alpha_i + \beta_i x_{ij},$$

the term $\log|\ell_{\alpha_i;\hat{\alpha}_i}(\beta_i, \hat{\alpha}_{\beta_i}; \hat{\beta}_i, \hat{\alpha}_i, A_i)|$ depends on data only, where $\hat{\alpha}_{\beta_i}$ is the ML estimator of α_i when the value of β_i is treated as fixed, so that we have an explicit form of modified

profile likelihood; up to a constant,

$$\ell_i^M(\beta_i) = \ell_i^P(\beta_i) + \frac{1}{2}\log|j_{\alpha_i\alpha_i}(\beta_i, \hat{\alpha}_{\beta_i}; \hat{\beta}_i, \hat{\alpha}_i, A_i)| \quad (5)$$

$$= \sum_j \{z_{ij}\log(\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}}) + m_{ij}\log(1-\hat{\pi}_{ij})\} + \frac{1}{2}\log\sum_j \{m_{ij}\hat{\pi}_{ij}(1-\hat{\pi}_{ij})\}, \quad (6)$$

where $\hat{\pi}_{ij}$ is given by $1/(1+\exp(-\hat{\alpha}_{\beta_i} - \beta_i x_{ij}))$. In this paper we show that $\ell_i^M(\beta_i)$ is an accurate approximation for $f(\boldsymbol{z}_i|T_i = t_i; \beta_i)$ in (4), but have no computational difficult.

2.3 Conditional likelihood for a random effect model

2.3.1 Laplace approximation

Due to the nature of the way collecting data in meta-analysis and other fields, we usually need to allow for heterogeneity among different data sets. For this a random effect model is often proposed to use; for example

$$\beta_i \sim N(\mu_\beta, \tau^2). \quad (7)$$

The full conditional likelihood for unknown parameters (μ_β, τ^2) is given by

$$l(\mu_\beta, \tau^2) = \sum_{i=1}^K \log\int_\beta f(\boldsymbol{z}_i|T_i = t_i; \beta)\phi(\beta; \mu_\beta, \tau^2)d\beta \quad (8)$$

where $f(\boldsymbol{z}_i|T_i = t_i; \beta)$ is given by (4), and $\phi(\beta; \mu_\beta, \tau^2)$ is the density function of the normal distribution $N(\mu_\beta, \tau^2)$. However, an explicit form of this is not often feasible, so that we may either use a numerical method such as Gauss-Hermite quadratures or the Laplace approximation. Tierney and Kadane (1986) showed that Laplace method is often sufficiently accurate and easy to obtain. For binary generalized linear mixed models (GLMMs), Noh and Lee (2007) showed that Laplace approximation is accurate enough in practice. It can be used even when the dimensionality of integral is large, while Gauss-

Hermite quadratures cannot. In this paper we use the Laplace approximation following Lee, Nelder and Pawitan (2006,Ch.4). Gauss-Hermite quadratures can be used when the dimensionality of integral is not large.

In this paper, we propose the following h-likelihood inferential procedure.

- Define the h-likelihood using the modified-profile likelihood

$$h^M = \sum_i h_i^M$$

  where $h_i^M = \ell_i^M(\beta_i) + \log \phi(\beta_i; \mu_\beta, \tau^2)$, $\ell_i^M(\beta_i)$ is the modified profile likelihood (5).

- For estimating mean parameter μ_β, use the Laplace approximation $p_\beta(h^M)$ to the marginal likelihood

$$p_\beta(h^M) = h^M(\hat{\beta}) - \frac{1}{2}\log\left|(-\partial^2 h^M/\partial\beta\partial\beta^t)/(2\pi)\right|_{\beta=\hat{\beta}},$$

  where $\hat{\beta}$ =argmax$_\beta h(\beta)$

- For estimating dispersion parameter τ^2, use $p_\delta(h^M)$ with $\delta = (\beta, \mu_\beta)$

Lee and Nelder (1996) proposed the use of h-likelihood for inferences from hierarchical generalized linear models. They define $p_\delta(h)$ as an approximation to the restricted likelihood (Lee, Nelder and Pawitan, 2006, Section 5.2.2). In this paper we show the h-likelihood approach gives practically useful inferences for the meta analysis, by comparing MCMC-EM algorithm of Shi and Copas (2002).

2.3.2 Markov chain Monte Carlo EM algorithm

Shi and Copas (2002) proposed the use of MCMC-EM algorithm for meta analysis. The EM algorithm arises because the true value of β_i can be thought of as a missing observa-

tion, and the MCMC algorithm arises because the E-step of the EM algorithm involves an integral with respect to the distribution of β_i. They used the Metropolis-Hastings algorithm to generate a Markov chain of β_i in the E-step of every iteration, such that its equilibrium distribution is the conditional distribution of β_i given the current estimates of unknown parameters.

We now extend the MCMC-EM algorithm to the above random effect model by using conditional likelihood approach. Because of difficulty in computing the exact conditional likelihood we propose to use the modified profile likelihood as follows:

- E-step: In the $(r+1)$th iteration, we calculate the conditional expectation

$$E(\sum_i \{\ell_i^M(\beta_i) + \log \phi(\beta_i; \mu_\beta, \tau^2)\} | \boldsymbol{z}, \mu_\beta^{(r)}, \tau^{2(r)})$$

  In this and subsequent expressions, expectations are also conditional on $T_i = \sum_j z_{ij} = t_i$. As there is no analytical form for the above equation, we use MCMC algorithm to calculate it.

  The Metropolis-Hastings Algorithm is used to generate a random variate β_i from its conditional distribution.

$$p\left(\beta_i | z_i, T_i = t_i\right) \propto \exp(\ell_i^M(\beta_i))\phi(\beta_i; \mu_\beta, \tau^2).$$

  Suppose that β_i^a is the random variate generated at the ath iteration. Then at the $(a+1)$th iteration we generate a random number β_i from an aperiodic recurrent transition density $q(\beta_i^a, \beta_i)$ and accept it as β_i^{a+1} with acceptance probability

$$\min\left(1, \frac{p(\beta_i | z_i, T_i = t_i) q(\beta_i, \beta_i^a)}{p(\beta_i^a | z_i, T_i = t_i) q(\beta_i^a, \beta_i)}\right)$$

- M-step: We get the explicit formula for μ_β, τ^2.

$$\begin{aligned}\hat{\mu}_\beta &= \frac{1}{K}\sum_{i=1}^{K} E(\beta_i|z_i, T_i = t_i) \\ \hat{\tau^2} &= \frac{1}{K}\sum_{i=1}^{K} E((\beta_i - \bar{\beta})^2|z_i, T_i = t_i).\end{aligned}$$

  Let $\{\beta_i^a, a = 1, \ldots, A\}$ are the random variates generated in the MCMC E-step, and let

$$\hat{\mu}_\beta^a = \frac{1}{K}\sum_{i=1}^{K}\beta_i^a, \quad \hat{v}^a = \frac{1}{K}\sum_{i=1}^{K}(\beta_i^a - \hat{\mu}_\beta^a)^2$$

  then μ_β and τ^2 can be updated by the sample means of $\hat{\mu}_\beta^a$'s and $\hat{v}^a$'s, respectively.

Bellio and Sartori (2003) showed that the modified profile likelihood gives an accurate approximation to the conditional likelihood when $m_{ij} \geq 5$. Furthermore, it is numerically much faster than by using the exact density because it does not have combination terms.

We discussed conditional inference for mixed effects models with binary data in this section by using h-likelihood Laplace approximation method and MCMC-EM algorithm. Although the derivation is for one dimensional random effect only (i.e., β is univariate), there is no difficult to extend the methods to the model with high-dimensional random effects, see for example the meta-analysis model for multi-arm trials discussed in Section 4.

3 Meta-analysis and trend estimation

3.1 Conditional inference for meta-analysis with selection bias

We now apply the proposed approaches to meta-analysis and trend estimation with selection bias (see the details in Shi and Copas, 2004). The meta-analysis model is given in

(1) to (3), where the ith data set is the data collected in the ith study. The parameter β_i measures the association between the log-odds ratio and the dose level x_{ij}. Shi and Copas (2004) used a normal approximated model based on empirical logistic transformation which is not suitable for small m_{ij} as discussed in Section 1.

To address the problem of publication bias, a selection model is defined (see Shi and Copas, 2002 and 2004). Let $\hat{\beta}$ be the estimate of slope in an individual study and s is its standard error and $\mathcal{S}$ be the event that a study is selected. To model the possibility that the selection is biased in favour of larger studies (with smaller value of s_i), and in favour of studies having a more positive outcome (with larger value of $\hat{\beta}_i$), suppose that a study reporting estimate $\hat{\beta}_i$ and standard error s_i is selected with probability

$$q(\boldsymbol{z_i}|\beta_i) = P(\mathcal{S}_i|\hat{\beta}_i, s_i, \beta_i) = \Phi\left(\frac{a + b/s_i + \rho(\hat{\beta}_i - \beta_i)/s_i}{(1-\rho^2)^{1/2}}\right), \tag{9}$$

where $b \geq 0,\ \rho \geq 0$, Φ is the cumulate distribution function of the standard normal distribution. Copas and Shi (2000) argued that a and b are not estimable without making strong assumptions. They proposed a sensitivity analysis instead: give a range of different values of (a, b) and then monitor how sensitively the estimate (μ_β, τ, ρ) and other quantities depend on the particular choice of these selection parameters.

For the meta-analysis of K studies, the log-likelihood is for those selected studies, and it is therefore given by

$$\begin{aligned}
\ell(\boldsymbol{Z};\boldsymbol{\theta}) &= \sum_{i=1}^{K} \log\{p(\boldsymbol{z}_i|T_i = t_i, \mathcal{S}_i, \boldsymbol{\theta})\} \\
&= \sum_{i=1}^{K} \log\left\{\frac{p(\boldsymbol{z}_i, \mathcal{S}_i|T_i = t_i, \boldsymbol{\theta})}{p(\mathcal{S}_i)}\right\} \\
&= \sum_{i=1}^{K} \{\log p(\boldsymbol{z}_i, \mathcal{S}_i|\mathcal{C}, \boldsymbol{\theta}) - \log p(\mathcal{S}_i)\}
\end{aligned}$$

$$= \sum_{i=1}^{K} \left\{ \log \int_{\beta_i} f(\boldsymbol{z}_i|T_i = t_i; \beta_i) q(\boldsymbol{z}_i|\beta_i) \phi(\beta_i : \mu_\beta, \tau^2) d\beta_i - \log p(\mathcal{S}_i) \right\} \quad (10)$$

$$\simeq \sum_{i=1}^{K} \left\{ \log \int_{\beta_i} \exp(\ell_i^M(\beta_i)) q(\boldsymbol{z}_i|\beta_i) \phi(\beta_i : \mu_\beta, \tau^2) d\beta_i - \log p(\mathcal{S}_i) \right\} \quad (11)$$

We need to consider the problems similar to (8), but the integrand is more complicated than the one in (8). In (11), $\log p(\mathcal{S}_i)$ could be ignored.

Here we continue to use the modified profile likelihood $\ell_i^M(\beta_i)$ as an approximation to $\log f(\boldsymbol{z}_i|T_i = t_i; \beta_i)$ in (10). Even though the model become more complicated due to the inclusion of selection model, the implementation algorithm remains essentially the same. The proposed h-likelihood inferential procedure is as follows.

- Define the (modified) h-likelihood

  $$h^M = \sum_i h_i^M$$

  where $h_i^M = \ell_i^M(\beta_i) + \log q(z_i|\beta_i) + \log \phi(\beta_i; \mu_\beta, \tau^2)$.

- For estimating mean parameter μ_β, use the Laplace approximation $p_\beta(h^M)$ to the marginal likelihood

  $$p_\beta(h^M) = h^M(\hat{\beta}) - \frac{1}{2} \log \left| (-\partial^2 h^M / \partial\beta \partial\beta^t)/(2\pi) \right| |_{\beta=\hat{\beta}},$$

  where $\hat{\beta}$ =argmax$_\beta h(\beta)$

- For estimating dispersion parameter τ^2 and selection parameter ρ, use $p_\delta(h^M)$ with $\delta = (\beta, \mu_\beta)$

3.2 Trend estimation for alcohol use and breast cancer

To study the association between breast cancer and alcohol consumption, a number of epidemiologic investigations have been conducted. Table 1 reports the results for such a study (Hiatt and Bawol, 1984). In this follow-up study, each row is correspond to an exposure band, including the base-line group with zero dose. The empirical odds ratio reported in the last column gives an estimate of the odds of being a case versus being a control. We apply the model defined in Section 2.1 but for this individual data-set only, and obtain the estimates $\hat{\alpha}_1 = -4.6117$ and $\hat{\beta}_1 = 0.0092$. The value of β measures the relation between the alcohol consumption and the risk of breast cancer. The estimate $\hat{\beta}_1 = 0.0092$ implies that one extra drink daily (about 13 gram of alcohol) increases risk by about 12 per cent.

There are total 14 studies used in a meta-analysis. Since some of the studies are case-control and some are follow-up, the values of α_i are quite different, and it is impossible to assume any distribution for α_i. The numerical study shows that the calculation of the estimates of α_i are very unstable by using unconditional approach and the calculation of the estimate β is also not reliable. Both of them are very sensitive to the choice of starting values in the iteration process.

In the contrast, there is only three parameters, $\theta = (\mu_\beta, \tau^2, \rho)$, are involved in the model by using conditional approach. The calculation of the estimates are very stable, and converges very fast. As an illustration, we consider the models with the following three typical pairs of (a, b) in sensitivity analysis (see the details in Shi and Copas, 2004):

$$(0.9292, 0.0017), \quad (-1.4061, 0.0073), \text{ and } (-2.4769, 0.0056).$$

We used both the h-likelihood approach based upon Laplace approximation and MCMC-EM algorithm discussed in the previous section. The results are given in Table 2. Both methods give similar analysis. MCMC-EM is computationally intensive, and choice of sample size and stopping rule are rather subjective, while the h-likelihood procedure is computationally straightforward, using Newton-Rhapson method. For the case without selection bias, we observed that the h-likelihood procedure takes 6 seconds, while MCMC-EM algorithm takes 186 seconds.

4 Meta-analysis for multi-arm trials

4.1 Conditional likelihood

Consider a model for multi-arm trials of Lu and Ades (2004)

$$z_{ij} \sim Bin(m_{ij}, \pi_{ij}), \tag{12}$$

where $i = 1, \ldots, K$ and $j = 1, \ldots, J$. Here, we assume that there are J treatments involved in the trials and there are K studies in a meta-analysis. Choosing treatment 1 as the baseline, and θ_{ij} as the treatment effect of j-th treatment relative to the baseline, we can define a random effect model for multi-arm trials as follows

$$\begin{aligned}
\text{logit}(\pi_{i1}) &= \mu_i - \theta_{i2}/J - \theta_{i3}/J - \ldots - \theta_{iJ}/J \\
\text{logit}(\pi_{i2}) &= \mu_i + (J-1)\theta_{i2}/J - \theta_{i3}/J - \ldots - \theta_{iJ}/J \\
&\ldots \\
\text{logit}(\pi_{iJ}) &= \mu_i - \theta_{i2}/J - \theta_{i3}/J - \ldots + (J-1)\theta_{iJ}/J
\end{aligned} \tag{13}$$

where $\text{logit}(\pi) = \log \pi/(1-\pi)$, and

$$\boldsymbol{\theta}_i = (\theta_{i2}, \ldots, \theta_{iJ}) \sim N(\boldsymbol{\eta}, \boldsymbol{\Sigma}), \tag{14}$$

where $\boldsymbol{\eta} = (\eta_2, \ldots, \eta_J)$ are $(J-1)$-dimensional overall log-odds ratio for j-th treatment $(j = 2, \ldots, J)$ relative to the baseline (treatment 1), and $\boldsymbol{\Sigma}$ is the related covariance matrix. Lu and Ades (2004) suggested two special forms, which are

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdot & \rho \\ \rho & 1 & \cdot & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \tag{15}$$

or

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_2^2 & \rho\sigma_2\sigma_3 & \cdot & \rho\sigma_2\sigma_J \\ \rho\sigma_2\sigma_3 & \sigma_3^2 & \cdot & \rho\sigma_3\sigma_J \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_2\sigma_J & \rho\sigma_3\sigma_J & \cdots & \sigma_J^2 \end{bmatrix} \tag{16}$$

with $\rho = 1/2$. Here, σ^2 or $\{\sigma_j, j = 2, \ldots, J\}$ are unknown parameters. Let

$$\boldsymbol{X} = \begin{bmatrix} -1/J & -1/J & \cdot & -1/J \\ (J-1)/J & -1/J & \cdot & -1/J \\ -1/J & (J-1)/J & \cdot & -1/J \\ \vdots & \vdots & \ddots & \vdots \\ -1/J & -1/J & \cdot & (J-1)/J \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1^T \\ \boldsymbol{X}_2^T \\ \boldsymbol{X}_3^T \\ \cdots \\ \boldsymbol{X}_J^T \end{bmatrix},$$

the model (13) can be rewritten as

$$\eta_{ij} = \text{logit}(\pi_{ij}) = \mu_i + \boldsymbol{X}_j^T \boldsymbol{\theta}_i. \tag{17}$$

By the same reasons discussed before, we use a conditional likelihood to eliminate the nuisance parameters $\{\mu_i, i = 1, \ldots, K\}$ in the above equations.

In meta-analysis, some studies may include only part of those J treatments. Let I_i be the indices of the treatments involved in the i-th study. Thus the data involved in the i-th study is

$$\mathcal{D}_i = \{(z_{ij}, m_{ij}), j \in I_i\}.$$

The conditional likelihood for i-th study is therefore given by

$$f(\mathcal{D}_i|\sum_{j\in I_i} z_{ij}=t_i,\boldsymbol{\theta}_i)=\frac{\prod_{j\in I_i}\begin{pmatrix} m_{ij} \\ z_{ij}\end{pmatrix}\exp(z_{ij}\boldsymbol{X}_j^T\boldsymbol{\theta}_i)}{\sum_{\boldsymbol{u}\in\mathcal{U}_{ij}}\prod_{j\in I_i}\begin{pmatrix} m_{ij} \\ u_j\end{pmatrix}\exp(u_j\boldsymbol{X}_j^T\boldsymbol{\theta}_i)}, \tag{18}$$

where $\boldsymbol{u}=(u_1,\cdots,u_J)'$, and

$$\mathcal{U}_{ij}=\{\boldsymbol{u}: 0\le u_j\le m_{ij}, \text{ and } \sum_{j\in I_i} u_j=t_i\}.$$

The full conditional likelihood for unknown parameters involved in $(\boldsymbol{\eta},\boldsymbol{\Sigma})$ is

$$l(\boldsymbol{\eta},\boldsymbol{\Sigma})=\sum_{i=1}^{K}\log\int_{\boldsymbol{\theta}_i} f(\mathcal{D}_i|\sum_{j\in I_i} z_{ij}=t_i,\boldsymbol{\theta}_i)\phi(\boldsymbol{\theta}_i;\boldsymbol{\eta},\boldsymbol{\Sigma})d\boldsymbol{\theta}_i \tag{19}$$

$$\simeq\sum_{i=1}^{K}\log\int_{\boldsymbol{\theta}_i}\exp(\ell_i^M(\boldsymbol{\theta}_i))\phi(\boldsymbol{\theta}_i;\boldsymbol{\eta},\boldsymbol{\Sigma})d\boldsymbol{\theta}_i, \tag{20}$$

where $\ell_i^M(\boldsymbol{\theta}_i)$ is the modified profile likelihood (6), approximating $\log f(\mathcal{D}_i|\sum_{j\in I_i} z_{ij})$, $\phi(\cdot;\boldsymbol{\eta},\boldsymbol{\Sigma})$ is the pdf of the multivariate normal distribution (14) with covariance matrix (15) or (16).

For the analysis of the data in multi-arm trials the equation (8) involves a $(J-1)$-dimensional integration. With the numerical method such as Gaussian-Hermite quadratures methods the calculation becomes much more difficult as J increases. In the h-likelihood approach, the implementation algorithm remains essentially the same. The proposed h-likelihood inferential procedure is as follows.

- Define the (modified) h-likelihood

$$h^M=\sum_i h_i^M$$

  where $h_i^M=\ell_i^M(\boldsymbol{\theta}_i)+\log\phi(\boldsymbol{\theta}_i;\boldsymbol{\eta},\boldsymbol{\Sigma})$.

- For estimating mean parameter $\boldsymbol{\eta}$, use the Laplace approximation $p_{\boldsymbol{\theta}}(h^M)$ to the

marginal likelihood where $\boldsymbol{\eta} = (\boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_K)$

$$p_{\boldsymbol{\theta}}(h^M) = h^M(\hat{\boldsymbol{\theta}}) - \frac{1}{2}\log\left|(-\partial^2 h^M/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t)/(2\pi)\right|\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $\hat{\boldsymbol{\theta}}$ =argmax$_{\boldsymbol{\theta}} h^M(\boldsymbol{\theta})$

- For estimating covariance parameters involved in $\boldsymbol{\Sigma}$, use $p_\delta(h^M)$ with $\delta = (\boldsymbol{\theta}, \boldsymbol{\eta})$

4.2 Aspirin for preventing vascular event

In total there are 39 randomised trials investigating the use of medium dose aspirin and/or high dose aspirin in the prevention of vascular events in high risk patients, making the comparison between three treatments High dose(500-1500 mg/day) aspirin(A), Medium dose(75-325 mg/day) aspirin(B) and control (C), where 3 trials compare A and B, 19 trials compare A and C, and 17 trials compare B and C (see the details in Song, *et al.*, 2003). The data are given in Table 3.

To use the model given in the previous section, the control group is treated as treatment 1, and the high dose (A) and medium dose (B) are treated as treatment 2 and 3 respectively. The first number in the table is z_{ij} and the second number is m_{ij} in (12). The first three studies include treatments 2 and 3 only, studies 4 to 22 include treatment 1 and 2 only while studies 23 to 39 include treatment 1 and 3 only. Thus

$$I_i = \begin{cases} \{2,3\} & \text{for } i = 1,2,3 \\ \{1,2\} & \text{for } i = 4,\ldots,22 \\ \{1,3\} & \text{for } i = 23,\ldots,39 \end{cases}$$

The estimates by using conditional approach are given in Table 4. The estimates by both methods are quite close. Due to the slow convergence of Metropolis-Hastings algorithm, the computational burden for MCMC-EM algorithm is much heavier than the h-likelihood

Laplace approximation method.

5 Simulation Study

We conduct a simulation study in this section to study the performance of the proposed methods and compare the conditional and unconditional approaches. Numerical studies, based upon 100 replications of simulated data, are presented in Table 5. We only consider the hierarchical likelihood approach based on Laplace approximation because MCMC-EM method is computationally too intensive. First we consider the single random effect problem of alcohol use and breast cancer data in Section 3.2. Simulated data are generated from the following logistic model,

$$\log(\pi_{ij}/(1-\pi_{ij})) = \alpha_i + \beta_i x_{ij}$$

where α_i and β_i are sampled from models

$$\text{M1} \quad : \quad \alpha_i \sim N(-2, 1^2) \ \text{ and } \ \beta_i \sim N(\mu_\beta, \tau^2)$$

$$\text{M2} \quad : \quad \alpha_i \sim 0.6 * N(-2, 1^2) + 0.4 * N(1, 1^2) \ \text{ and } \ \beta_i \sim N(\mu_\beta, \tau^2).$$

We use the same covariates and binomial denominators of the original dataset. Three pairs of values for (μ_β, τ^2) are considered (see the second column in Table 5). In each replication, we generated data based on the above real values of the parameters, and then calculate the estimates by using conditional likelihood approach based on h-likelihood Laplace approximation. We report the mean and standard error for the estimates of (μ_β, τ^2) in Table 5, in which $\hat{\mu}_\beta$ and its standard error are defined by $\sum_i \hat{\mu}^i_\beta/100$ and $\sqrt{\sum_i (\hat{\mu}^i_\beta - \hat{\mu}_\beta)^2/(99 \times 100)}$ respectively based on 100 replications. Note that $\hat{\mu}^i_\beta$ is the

estimate of μ_β in the ith replication and $\hat{\mu}_\beta$ is the sample mean of $\hat{\mu}_\beta^i$'s. The values of $\hat{\tau}^2$ and its standard error are calculated similarly. We used R in the simulation study. Table 5 shows that the proposed method give very accurate results for both simulation models M1 and M2.

As comparison, we also used the unconditional likelihood approach. The results are given in Table 6. This approach has many nuisance parameters α_i, which make the convergence to be slow, having occasional divergences (see the numbers listed in the last column). Comparing with the conditional approach, although the mean of $\hat{\mu}_\beta$'s by using unconditional approach is also quite close to the true value, but the estimates is much more variable. Because of increase in the number of nuisance parameters the dispersion parameter τ^2 is often under-estimated.

Secondly, we consider the multi-dimensional random effects problem of Aspirin for preventing vascular event in Section 4.2. We did not consider the unconditional approach because the lager number of divergences. Simulated data are generated using the logistic model,

$$\log(\pi_{ij}/(1-\pi_{ij})) = \mu_i + X_j^T \theta_i$$

where μ_i and θ_i are from models

$$\text{M3} \quad : \quad \mu_i \sim N(-2, 0.1^2) \ \text{ and } \ \theta_i \sim BVN((\eta_2, \eta_3), \mathbf{\Sigma})$$

$$\text{M4} \quad : \quad \mu_i \sim 0.6 * N(-2, 0.1^2) + 0.4 * N(1, 0.1^2) \ \text{ and } \ \theta_i \sim BVN((\eta_2, \eta_3), \mathbf{\Sigma})$$

and

$$\mathbf{\Sigma} = \tau^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

and $BVN(0, \mathbf{\Sigma})$ means bivariate normal distribution with mean 0, covariance $\mathbf{\Sigma}$. We also use the same covariates and binomial denominators of the original dataset. For the above simulation models, we conducted a simulation study by using conditional approach with h-likelihood Laplace approximation. The simulation results are reported in Table 7. As before, the true values of the parameters are given in the second column, and the mean and standard error of the related estimates are given in the other columns. The proposed method gives very accurate results for both simulation models M3 and M4.

6 Discussion

We discussed a hierarchical logistic regression model with mixed effects for binary data in this paper. Conditional likelihood approach is a sensible choice to eliminate the many nuisance parameters and give consistent estimates for the parameters of interest. However, as we have discussed, the heavy computational burden limits the use of the approach. We proposed h-likelihood Laplace approximation method and MCMC-EM algorithm in this paper. Both methods give accurate results. The computation for the former is very straightforward, although the MCMC-EM method is still quite computational intensive especially for the model with multi-dimensional random effects.

We applied the methods to two important models. One is the meta-analysis model with publication bias for trend estimation. The inference based on conditional likelihood and the exact binomial distribution is difficult, and the inference is more intractable for the model with publication bias. However, both proposed methods perform very well and are computationally efficient. The second application is for the model with multi-arm trials. This is the model having multi-dimensional random effects, and the computa-

tion based on conditional likelihood is extremely intractable. The h-likelihood Laplace approximation method still works very efficiently, even for the model with large dimensional random effects. There is rich literature in medical science and other areas on using similar hierarchical or multi-level mixed effects models. We use those two typical examples to demonstrate how the two proposed methods work based on conditional likelihood approach.

Although the derivation given in this paper is for hierarchical logistic regression model only, there is no major difficult to extend the proposed methods to other types of generalized linear regression models with mixed effects.

Appendix: Proof of equation (5)

Since $Z_{ij}|\beta_i \sim Bin(m_{ij}, \pi_{ij})$ where $\log(\pi_{ij}/(1-\pi_{ij})) = \alpha_i + \beta_i x_{ij}, \quad j = 0, 1, ..., n_i$. The log likelihood $\ell_i(\beta_i, \alpha_i; \hat{\beta}_i, \hat{\alpha}_i, A_i)$ in Section 2.2 is

$$\sum_j [\log \begin{pmatrix} m_{ij} \\ z_{ij} \end{pmatrix} + z_{ij}(\alpha_i + \beta_i x_{ij}) + m_{ij} \log(1 - \pi_{ij})]$$

where $z_{ij} = z_{ij}(\hat{\beta}_i, \hat{\alpha}_i, A_i)$. Let $\hat{\beta}_i$ and $\hat{\alpha}_i$ be parameter estimates, only depending upon the data. Note that

$$\partial \ell_i(\beta_i, \alpha_i; \hat{\beta}_i, \hat{\alpha}_i, A_i)/\partial \alpha_i = \sum_j (z_{ij} - m_{ij}\pi_{ij}),$$

so that

$$\ell_{\alpha_i;\hat{\alpha}_i}(\beta_i, \hat{\alpha}_{\beta_i}; \hat{\beta}_i, \hat{\alpha}_i, A_i) = \partial^2 \ell_i(\beta_i, \alpha_i; \hat{\beta}_i, \hat{\alpha}_i, A_i)/\partial \alpha_i \partial \hat{\alpha}_i|_{\alpha_i = \hat{\alpha}_{\beta_i}}$$

depends on the data only because $\partial \sum_j (m_{ij}\pi_{ij})/\partial \hat{\alpha}_i = 0$. Thus, we can ignore $\ell_{\alpha_i;\hat{\alpha}_i}(\beta_i, \hat{\alpha}_{\beta_i}; \hat{\beta}_i, \hat{\alpha}_i, A_i)$ to give the modified profile likelihood (5).

References

Andersen, EB. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. R. Statist. Soc.* B, **32**, 283-301.

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. R. Statist. Soc.* B, **41**, 279-312.

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

Bellio, R. and Sartori, N. (2006). Practical use of modified maximum likelihoods for stratified data, *Biometrical Journal*, **48**,876-886.

Bellio, R. and Sartori, N. (2003). Extending conditional likelihood in models for stratified binary data, *Statistical Methods & Applications*, **12**,121-132.

Copas, J.B. and Shi, J.Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, **1**, 247-262.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *J. R. Statist. Soc.* B, **49**, 1-39.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*. (2nd edition). Chapman and Hall.

Davison, A. C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc.* B, **50** , 445-461.

Hiatt, R. A. and Bawol, R. D. (1984). Alcoholic beverage consumption and breast cancer incidence. Am. J. Epidemiol, 120, 676-683.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with

discussion). *J. R. Statist. Soc.* B, **58**, 619-656.

Lee, Y. and Nelder, J.A. (2002). Analysis of ulcer data using hierarchical generalized linear models. *Statistics in Medicine*, **21**,191-202.

Lee, Y., Nelder, J.A. and Pawitan, Y.(2006). *Generalized Linear Models with Random Effects*. Chapman and Hall.

Liao, J. G. (1999) A Hierarchical Bayesian model for combining multiple 2×2 tables using conditional likelihoods. *Biometrics*, **55**, 268-72.

Lu, G. and Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, **23**, 3105-3124.

Lubin, J. H. (1981). An empirical evaluation of the use of conditional and unconditional likelihoods for case-control data. *Biometrika*, **68**, 567-71.

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a neo-Fisherian perspective*. Singapore: World Scientific Publishing.

Noh, M. and Lee, Y. (2007). REML estimation for binary data in GLMMs, *J. of Multivariate Analysis*, **98**, 896-915.

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters, *Biometrika*, **90**,533-549.

Severini, T.A. (2000). *Likelihood Methods in Statistics*. Oxford:Oxford University Press.

Shi, J. Q. and Copas, J. B. (2002). Publication bias and meta-analysis for 2×2 tables: an average Markov chain Monte Carlo EM algorithm. *J. R. Statist. Soc.* B, **64**, 221-236.

Shi, J. Q. and Copas, J. B. (2004). Meta-analysis for trend estimation. *Sta-*

tistics in Medicine, **23**, 3-19.

Song, F., Altman, D. G., Glenny, A. M. and Deeks, J. J. (2003). Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published metaanalyses. *British Medical Journal*, **326**. 472-6.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *JASA*, **81**, 82-6.

Vollset, S. E., Hirji, K. F. and Elashoff, R. M. (1991). Fast computation of exact confidence limits for the common odds ratio in a series of 2×2 tables. *JASA*, **86**, 404-9.

Table 1. Follow-up data on alcohol use and breast cancer

Alcohol (g/day)	Assigned dose x	No. of cases	No. of controls	Empirical OR
0	0	252	24089	1.0
< 26	6.8	505	49432	1.024
39-65	46.34	68	3892	1.670
> 78	83.6	13	760	1.635

Table2. . Result of Meta analysis for each (a, b)

Methods	(a, b)	μ_β (s.e.)	τ^2 (s.e.)	ρ (s.e.)
H[1]	$(0.9292, 0.0017)$	0.009702 (0.000968)	0.00001311 (0.000006383)	0.9900 (fixed[3])
EM[2]	$(0.9292, 0.0017)$	0.009211 (0.001135)	0.00001720 (0.000006409)	0.9792 (0.0430)
H	$(-1.4061, 0.0073)$	0.008179 (0.001084)	0.00001646 (0.000007531)	0.6827 (0.2560)
EM	$(-1.4061, 0.0073)$	0.007854 (0.001080)	0.00001641 (0.000006136)	0.7158 (0.1642)
H	$(-2.4769, 0.0056)$	0.006235 (0.001080)	0.00001634 (0.000007565)	0.5284 (0.1376)
EM	$(-2.4769, 0.0056)$	0.005638 (0.001158)	0.00001845 (0.000007192)	0.6093 (0.1064)
H	without assuming	0.009965 (0.001030)	0.00001486 (0.000007289)	
EM	selection bias	0.009890 (0.001031)	0.00001495 (0.000006083)	

1. H means hierarchical likelihood approach based on Laplace approximation.
2. EM means MCMC-EM algorithm.
3. A value of ρ close to 1 is not realistic because it implies that accepting a study for review is then simply a matter of comparing the empirical log odds with a fixed threshold. Following Shi and Copas (2002), we assume $\rho \leq 0.99$.

Table 3. Randomized trials of Aspirin data

Study number	Number of patients		
	High dose(A) event/total	Medium dose(B) event/total	Control(C) event/total
1	168/815	174/806	
2	18/155	18/154	
3	9/242	9/253	
4	2/63		6/69
5	6/243		14/236
6	375/1856		406/1855
7	129/847		186/878
8	76/758		102/771
9	379/2267		411/2257
10	23/101		27/102
11	6/75		10/73
12	1/71		5/77
13	26/162		35/157
14	33/317		45/309
15	65/672		106/668
16	2/30		5/30
17	59/253		55/252
18	5/92		6/84
19	4/42		2/40
20	1/148		3/150
21	0/357		3/357
22	5/44		5/44
23		21/150	21/151
24		163/676	193/684
25		46/474	85/471
26		111/1009	159/1026
27		33/336	34/336
28		12/50	18/50
29		2/29	4/31
30		915/8587	1236/8600
31		1/19	1/25
32		33/313	46/306
33		5/28	1/28
34		57/615	76/624
35		124/404	127/378
36		1/26	4/24
37		45/661	75/677
38		49/552	49/568
39		5/37	5/33

Table 4. Result of meta-analysis with multi-arm trials

Methods	η_2 (s.e.)	η_3 (s.e.)	τ^2 (s.e.)
H	-0.2674 (0.01411)	-0.2956 (0.01411)	0.007766 (0.002615)
EM	-0.2585 (0.01218)	-0.2745 (0.01216)	0.005760 (0.001289)

Table 5. Simulation study: conditional analysis for single-arm trial

α	(μ_β, τ^2)	$\hat{\mu}_\beta$ (s.e.)	$\hat{\tau}^2$ (s.e.)
M1	$(0.01, 0.000015)$	0.01003 (0.00005976)	0.00001448 (0.0000004075)
	$(0.05, 0.000015)$	0.05005 (0.00007253)	0.00001384 (0.0000006807)
	$(0.01, 0.000050)$	0.009953 (0.0001828)	0.00004678 (0.000001992)
M2	$(0.01, 0.000015)$	0.01001 (0.00007797)	0.00001482 (0.0000007806)
	$(0.05, 0.000015)$	0.05008 (0.00007306)	0.00001386 (0.0000006145)
	$(0.01, 0.000050)$	0.01024 (0.0001636)	0.00004615 (0.000001901)

Table 6. Simulation study: unconditional analysis for single-arm trial

α	(μ_β, τ^2)	$\hat{\mu}_\beta$ (s.e.)	$\hat{\tau}^2$ (s.e.)	# of divergences
M1	$(0.01, 0.000015)$	0.009906 (0.0001425)	0.000008991 (0.0000009974)	3
	$(0.05, 0.000015)$	0.05003 (0.0001457)	0.000008375 (0.0000008892)	7
	$(0.01, 0.000050)$	0.009962 (0.0002679)	0.00003907 (0.000002104)	2
M2	$(0.01, 0.000015)$	0.01070 (0.0001503)	0.000009327(0.000009057)	6
	$(0.05, 0.000015)$	0.05010 (0.0001462)	0.000008531 (0.0000008992)	8
	$(0.01, 0.000050)$	0.01027 (0.0002318)	0.00003327(0.000002057)	6

Table 7. Simulation study: meta-analysis with multi-arm trials

	$(\theta_2, \theta_3, \tau^2)$	θ_2 (s.e.)	θ_3 (s.e.)	τ^2 (s.e.)
M3	$(-0.26, -0.29, 0.01)$	-0.2629 (0.004522)	-0.2884 (0.004437)	0.008378 (0.0009477)
M4	$(-0.26, -0.29, 0.01)$	-0.2685 (0.004183)	-0.2860 (0.004698)	0.008577 (0.0009514)