Meta-analysis for rare events

Tianxi Cai∗†, Layla Parast and Louise Ryan

Meta-analysis provides a useful framework for combining information across related studies and has been widely utilized to combine data from clinical studies in order to evaluate treatment efficacy. More recently, meta-analysis has also been used to assess drug safety. However, because adverse events are typically rare, standard methods may not work well in this setting. Most popular methods use fixed or random effects models to combine effect estimates obtained separately for each individual study. In the context of very rare outcomes, effect estimates from individual studies may be unstable or even undefined. We propose alternative approaches based on Poisson random effects models to make inference about the relative risk between two treatment groups. Simulation studies show that the proposed methods perform well when the underlying event rates are low. The methods are illustrated using data from a recent meta-analysis (N. Engl. J. Med. 2007; 356(24):2457–2471) of 48 comparative trials involving rosiglitazone, a type 2 diabetes drug, with respect to its possible cardiovascular toxicity. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: cardiac complications; rosiglitazone; Poisson random effects

1. Introduction

A number of recent high profile cases of adverse reactions associated with pharmaceutical products have heightened the need for careful pre- and post-marketing scrutiny of product safety. Since clinical trials are typically designed to assess efficacy, single studies rarely have sufficient power to provide reliable information regarding adverse effects, especially when these adverse events are rare. For this reason, it is natural to base safety assessments on meta-analyses that combine data from multiple related studies involving the same or similar treatment comparisons. For example, Nissen and Wolski [1] evaluated the effect of rosiglitazone, a type 2 diabetes drug, with respect to the risk of myocardial infarction (MI) and CVD-related death.

There is a vast literature on statistical methods for meta-analysis [2]. Though details vary, most popular approaches employ either fixed effects or random effects models to combine study-specific effect estimates such as the mean differences, relative risks, and odds ratios. A fixed effects approach assumes treatment effects to be the same across all study populations and commonly estimates this effect using a weighted average of study-specific point estimates. When data can be arrayed as a set of multiple $2 \times 2$ tables, the most commonly used procedures include the Mantel–Haenszel (MH) [3] and Peto methods [4]. When event rates are low, however, these standard methods may not perform well [5, 6]. Recently, Tian et al. [5] proposed an exact approach that combines study-specific exact confidence intervals to yield an overall interval estimation procedure for the fixed effects model.

In practice, the fixed effect assumption will generally be too restrictive. One remedy is to take random effects approaches that allow for between-study heterogeneity. Through a random effects framework, DerSimonian and Laird [7] developed useful procedures for estimating the mean of the random effect distribution. Their method was based on a linear combination of study-specific estimates of the parameter with the weights depending on the within- and between-study variation estimates. Though simple to implement, this method may not work well in the context of rare events where it is likely for several of the studies being combined to have no events at all. In the case of rosiglitazone, for example, 25 of the 48 studies include no deaths and 10 studies have no cardiac events. Nissen and Wolski took the conventional approach of omitting studies with zero events from their fixed effects odds ratio analysis. However, under the fixed effects framework, simulation results [5] indicate that the standard MH and Peto’s methods may lead to invalid or inefficient inference about the treatment difference when studies with zero events are deleted or artificially imputed. In this paper, we incorporate the zero event studies through a likelihood-based approach with a series of Poisson random
effects models for the number of adverse events. We show that leaving out studies with zero events may be justified using a conditional inference argument. However, simulation results suggest that such methods may not perform well when the study sample sizes are not sufficiently large. We also derive a more powerful procedure based on a hierarchical random effects model. Our proposed approaches extend those previously considered by Brumback et al. [8] and Gatsonis and Daniels [9]. We describe the general estimation procedure in Section 2 and the inference procedures in Section 3. In Section 4, we illustrate our proposed methods using data from the rosiglitazone study and present results from the numerical studies demonstrating that the methods work well in finite samples. Some concluding remarks are also given in Section 5.

2. A likelihood-based approach

Suppose we observe data from two treatment groups (e.g. treated versus control) in each of \( M \) different studies. Let treatment arm 1 be the treated group and treatment arm 2 be the control group. Let \( Y_{ij} \) be the number of adverse effects observed among the \( N_{ij} \) subjects in the \( j \)th treatment arm (\( j = 1, 2 \)) of the \( i \)th study (\( i = 1, \ldots, M \)). Let \( X_{ij} \) be a binary covariate indicating treatment group. Without loss of generality, let \( X_{ij} = 1 \) if treated group, 0 if control group. Since the data are counts, it is natural to assume that \( Y_{ij} \) follows a Poisson distribution with rate \( \lambda_{ij} \), where \( \lambda_{ij} \) follows a log-linear model. We begin by considering an oversimplified model assuming that all the different studies have the same baseline and relative risks:

\[
\lambda_{ij} = \bar{\lambda} e^{\tau X_{ij}}. \tag{1}
\]

The parameter \( \bar{\lambda} \) reflects the baseline event rate (corresponding to the control group with \( X_{ij} = 0 \)), while \( \tau \) represents the log-relative risk associated with treatment. To allow for between-study variation, it is desirable to vary both the baseline rates and log relative risks from study-to-study. We consider several approaches, starting out with simple models that assume a random baseline but a constant relative risk, then working up to models that also consider the \( \tau \) to vary from study-to-study. The former is relatively straightforward, while the latter may be rather complicated and several implementations will be considered.

2.1. Random baseline, fixed relative risk

Suppose we extend (1) to

\[
\lambda_{ij} = \bar{\xi}_i e^{\tau X_{ij}}, \tag{2}
\]

so that the baseline event rates vary from study-to-study, but the relative risk is assumed constant. The obvious approach of assuming a normal distribution on \( \log(\bar{\xi}_i) \) results in a non-closed form likelihood, necessitating the use of either numerical approaches such as quadrature or approximations based on Laplace transformations [10]. Alternatively, we may assume

\[
\bar{\xi}_i \sim \text{Gamma}(\alpha, \beta). \tag{3}
\]

Straightforward algebra establishes that the likelihood has a closed form and can be written as

\[
L_n(\theta) = \prod_{i=1}^{n} \frac{\beta^\alpha \Gamma(Y_i + \alpha) (N_{i1} e^\alpha + N_{i2})^{Y_i}}{\Gamma(\alpha) (\beta + N_{i1} e^\alpha + N_{i2})^{Y_i} \left[ Y_{i1}! \right]^{Y_{i1}} \left[ Y_{i2}! \right]^{Y_{i2}}},
\]

where \( Y_i = Y_{i1} + Y_{i2} \) is the total number of events observed in the \( i \)th study and \( \theta = (\alpha, \beta, \tau)' \) represents the parameters of interest. Brumback et al. [8] explored the use of this model to assess the impact of chorionic villus sampling on the occurrence of rare limb defects using data from a series of 29 studies. They showed that the likelihood (3) can be easily maximized using a non-linear numerical maximizer such as the function \texttt{nlmmin} in \texttt{R} and that the method works well for combining data on rare events from multiple studies, some of which have no events at all. The Poisson-gamma hierarchical framework has been used extensively in the literature for risk modeling [11–14].

2.2. Random relative risk

To allow for the possibility of study-to-study heterogeneity with respect to relative risk, we extend the models of Section 2.1 to

\[
\lambda_{ij} = \bar{\xi}_i \exp(\tau_i X_{ij}), \tag{4}
\]
where both the baseline parameters and the relative risks $\tau_i$ are assumed to be random. For example, the baseline risk $\xi_i \sim \text{Gamma}(z, \beta)$ and the log relative risk $\tau_i \sim N(\mu, \sigma^2)$. The resulting likelihood function is

$$
\prod_{i=1}^n \frac{\beta^2 \Gamma(Y_i + z)}{\Gamma(\beta + N_{i1}e^z + N_{i2})} \left(\frac{N_{i1}e^z}{Y_i!}\right)^{Y_i1} \left(\frac{N_{i2}}{Y_i2!}\right) \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\tau - \mu)^2}{2\sigma^2}} \, d\tau.
$$

When such approaches are taken, one may obtain approximations to the likelihood function based on procedures such as the Gaussian Quadrature, Laplace Approximation, or advanced simulation-based procedures such as the Markov Chain Monte Carlo (MCMC) methods [15]. While these procedures may work well for many settings, they tend to be computationally intensive and numerically unstable, especially in the context of rare events.

Alternatively, one may consider a fully Bayesian (FB) approach by specifying hyper-priors on the assumed random effects distributions. In addition to assuming gamma-distributed baseline risks, $\xi_i \sim \text{Gamma}(z, \beta)$, one may impose hyper-priors for both $z$ and $\beta$. The posterior distribution of $\tau_i$ may be obtained via MCMC methods. The FB approach, while useful, may be computationally intensive. In this section, we consider several alternative approaches.

2.2.1. Conditional inference. While random effects modeling provides an effective means of reducing the dimensionality associated with multiple study-specific parameters, an alternative classical dimension reduction approach is through conditioning on a sufficient statistic under the null hypothesis [16]. In the present case, the sufficient statistic for $\xi_i$ under $H_0: \tau_i = 0$ is $Y_i = Y_{i1} + Y_{i2}$. Hence we base inference on the conditional distribution of $Y_{i1}$ and $Y_{i2}$ given $Y_i$. It is straightforward to show that

$$
Y_{i1}|Y_i \sim \text{binomial}(Y_i, p_i),
$$

where $p_i = e^{z_i} / (W_i + e^{z_i})$ and $W_i = N_{i2}/N_{i1}$ is the ratio of sample sizes for the two treatment groups in the $i$th study. Simple algebra yields $\log(p_i) = \tau_i - \log(W_i)$, so that $\log(W_i)$ can be considered as a known offset in a logistic model on $p_i$. When $N_{i2} = N_{i1}$, it follows that $\tau_i$ is simply the logit of $p_i$, the probability that an adverse event observed in the $i$th study occurs in treatment arm 1. The null hypothesis of no treatment difference would correspond to $\tau_i = 0$ and $p_i = \frac{1}{2}$.

To model the between-study variability, one may impose a prior distribution on $\tau_i$, such as normal. However, the resulting likelihood function does not have a closed form, which would necessitate the use of quadrature or appropriate numerical approximations. We propose to instead model the $p_i$’s directly and assume that they follow an appropriate Beta distribution. In the absence of covariates, it is well known that the over-dispersed binomial data can often be conveniently characterized by a Beta–binomial distribution [17]. In the presence of covariates, the beta–binomial distribution can be extended by choosing the beta parameters so that the binomial response rates have, on average, the appropriate relationship with the covariates of interest, as well as the appropriate degree of study-to-study heterogeneity [9]. Here, we assume that $p_i \sim \text{Beta}(\psi \gamma, \psi W_i)$, with $\psi$ and $\gamma$ being the unknown parameters that control the center and the spread of the prior distribution. It is straightforward to see that

$$
E(p_i) = \mu_i = \frac{\gamma}{W_i + \gamma} \quad \text{and} \quad \text{var}(p_i) = \frac{\mu_i(1 - \mu_i)}{\psi(\gamma + W_i) + 1}.
$$

These moments have an appealing interpretation. The case with no treatment effect will be characterized by $\gamma = 1$ and a large value of $\psi$, so that the probability of an adverse event being on the treated arm will be purely driven by the relative sample sizes of the two arms. More generally, evidence of increased event rates on the treated arm will be characterized by values of $\gamma$ exceeding 1. The parameter $\psi$ serves to calibrate the amount of study-to-study variation in the treatment effects with a large $\psi$ essentially suggesting a fixed effects model.

To estimate the unknown parameters for the assumed Beta distribution, we note that the likelihood can be expressed as

$$
\hat{L}_n(\theta) = \prod_{i=1}^n p_i^{Y_{i1}}(1 - p_i)^{Y_{i2}} \frac{B(\psi \gamma + Y_{i1}, \psi W_i)}{B(\psi \gamma, \psi W_i)} \frac{B(\psi \gamma + Y_{i1} + Y_{i2}, \psi W_i)}{B(\psi \gamma, \psi W_i)},
$$

where $\theta = (\gamma, \psi)$. The closed form arises due to the conjugate relationship between the Beta and the binomial distributions. This likelihood can be easily maximized to estimate $\gamma$ and $\psi$ and standard likelihood theory can be applied to obtain confidence intervals or to construct hypothesis tests of the primary parameter of interest, $\gamma$.

2.2.2. Gamma distributed baseline risks. A disadvantage of the conditional inference approach is that only studies with at least one event can contribute to the analysis. Heuristically, it would seem that a more powerful analysis could be achieved through an unconditional random effects analysis that allows studies with zero events to contribute as well.
In this subsection, we consider an unconditional inference about model (4) assuming that baseline risks follow a gamma distribution with parameters $\alpha$ and $\beta$. In this case, the contribution of the $i$th study to the likelihood is:

$$L_i(\alpha, \beta; \tau_i) = \left\{ \frac{\beta^{Y_{i1}}(N_i)}{(\beta + N_i)^{Y_{i1}} \Gamma(Y_{i1} + 1) \Gamma(Y_{i2} + 1) \Gamma(\tau_i + 1/2)} \right\}^{Y_{i2}/2} \left( \frac{W_{i1} + e^{\tau_i}}{W_{i1} + e^{\tau_i} + 1} \right)^{n_{i1}} \left( \frac{W_{i2} + e^{\tau_i}}{W_{i2} + e^{\tau_i} + 1} \right)^{n_{i2}}$$

where $Y_i = Y_{i1} + Y_{i2}$, $p_{\beta i} = e^{\tau_i}/(W_{i1} + e^{\tau_i})$, and $W_{\beta i} = (N_i + e^{\tau_i})/N_i$. Similar to the conditional likelihood, the unconditional likelihood function may not have a closed form if a simple distributional assumption is imposed on $\tau_i$ directly. Since the first term in this re-expression has the kernel of a Beta distribution, we propose to place a Beta prior on the $p_{\beta i}$ and assume that $p_{\beta i} \sim \text{Beta}(\psi, \gamma,W_{\beta i})$, where $\beta$ is one of the parameters characterizing the distribution of the baseline rates and $\psi$ and $\gamma$ are unknown parameters that control the underlying distribution of the relative risk. From the properties of the Beta distribution, we have that

$$E(p_{\beta i}) = \mu_i = \frac{\gamma}{W_{\beta i} + \gamma} \text{ and } \text{var}(p_{\beta i}) = \frac{\mu_i(1 - \mu_i)}{\psi(\gamma + W_{\beta i} + 1)}.$$

Here, the parameter $\gamma$ controls the center of the distribution of $p_{\beta i}$ and the parameter $\psi$ drives the variance. If $\psi$ is very large, then there is essentially no variation in the $p_{\beta i}$'s. For a fixed value of $\beta$ and conditional on $W_i$, the log relative risk $\tau_i$ follows a logistic Beta distribution with an offset $\log(W_{\beta i})$.

The marginal likelihood can be obtained by integrating (6) over the Beta prior on the $p_{\beta i}$'s to obtain:

$$\hat{L}_n(\theta) = \prod_{i=1}^{n} \left\{ \frac{\beta^{Y_{i1}}(N_i)}{(\beta + N_i)^{Y_{i1}} \Gamma(Y_{i1} + 1) \Gamma(Y_{i2} + 1) \Gamma(\tau_i + 1/2)} \right\}^{Y_{i2}/2} \left( \frac{W_{i1} + e^{\tau_i}}{W_{i1} + e^{\tau_i} + 1} \right)^{n_{i1}} \left( \frac{W_{i2} + e^{\tau_i}}{W_{i2} + e^{\tau_i} + 1} \right)^{n_{i2}} B(\psi + Y_i, \psi W_{\beta i} + Y_{i2} + 1),$$

where $\theta = (\alpha, \beta, \psi, \gamma)'$ represents the expanded set of unknown model parameters.

All the methods described above are based on hierarchical models that include random effects characterized by distributions with unknown parameters. We take an empirical Bayes approach to inference, treating the parameters in the assumed random effects distributions as fixed quantities to be estimated along with the primary parameters of interest.

### 3. Inference procedures

For all the methods discussed in Section 2, we may obtain an estimator for the unknown parameter vector $\theta$ by maximizing the likelihood function with respect to $\theta$. Let $\hat{\theta} = \text{argmax}_{\theta} \hat{L}_n(\theta)$. Properties of the maximum likelihood estimators may be used to establish the consistency of $\hat{\theta}$ for the true hyper-parameter $\theta_0$ as well as the weak convergence of $n^{1/2}(\hat{\theta} - \theta_0) \rightarrow N(0, \Sigma)$, where $\Sigma = -\{E \hat{\theta}^2 \text{log} \hat{L}_n(\theta)/\text{d}\theta \hat{\theta} \}^{-1}$. The asymptotic covariance matrix may be estimated empirically using the observed information matrix. Interval estimates for $\theta$ or any component of $\theta$ may be constructed based on the normal approximation. Alternatively, one may construct confidence intervals for a specific component of $\theta$ based on the profile likelihood theory [18, 19]. For example, to construct confidence intervals for $\tau$ in (3), one may define the profile likelihood

$$\hat{L}_n^{\text{prof}}(\tau) = \text{sup}_{\alpha, \beta} \hat{L}_n(\alpha, \beta, \tau)$$

and put $\hat{\tau} = \text{argmax} \hat{L}_n^{\text{prof}}(\tau)$. An $\eta$-level confidence interval for $\tau$ can then be obtained as

$$\left\{ \tau: 2 \log \frac{\hat{L}_n^{\text{prof}}(\hat{\tau})}{\hat{L}_n^{\text{prof}}(\tau)} \leq d_\eta \right\}$$

where $d_\eta$ is the $\eta^{th}$ percentile of the $\chi^2$ distribution with one degree of freedom.

In addition to summarizing the relative risks based on the parameter $\theta$, one may also be interested in the mean relative risk across studies, i.e., $\mu_{RR} = E(e^{\tau_i})$, where the expectation is taken over the distribution of $\tau_i$ as well as the sample sizes across studies. For the conditional approach, an estimate of $\mu_{RR}$ may be obtained by $\hat{\mu}_{RR} = n^{-1} \sum_{i=1}^{n} f_{Ri}(\tau; \hat{\theta}) d\tau$, where

$$f_{Ri}(\tau; \theta) = \left( \frac{e^{\tau}}{W_i + e^{\tau}} \right)^{Y_i} \left( \frac{W_i}{W_i + e^{\tau}} \right)^{\psi W_i} B(\psi, \psi W_i).$$

Copyright © 2010 John Wiley & Sons, Ltd.
is the density of $\tau_i | W_i$. Under the Beta–Gamma framework, $\mu_{RR}$ can be estimated as 
$$
\hat{\mu}_{RR} = n^{-1} \sum_{i=1}^{n} \int e^\tau f_G(\tau; \hat{\theta}) d\tau,
$$
where

$$
f_G(\tau; \theta) = \left( \frac{e^\tau}{W_{\beta i} + e^\tau} \right)^{\psi_{\gamma}} \left( \frac{W_{\beta i}}{W_{\beta i} + e^\tau} \right)^{\psi_{W_{\beta i}}} / B(\psi_{\gamma}, \psi_{W_{\beta i}})
$$

is the density of $\tau_i | W_{\beta i}$. These density functions could also be used to assess the between-study variation in the relative risks such as var($e^{\hat{\tau}}$). Alternatively, one may also summarize the distribution of $\tau_i$ given a particular sample size allocation. For example, one might be interested in the extreme setting with $N_{i1} = N_{i2} \to \infty$. In that case, $W_i = 1$ and $W_{\beta i} \to 1$. The distribution of $\tau_i$ can be estimated accordingly.

It is important to note that both of the location parameter $\gamma$ and the scale parameter $\psi$ contains information on the treatment effect. To ascertain the treatment effects across studies, the entire distribution of $\tau_i$ should be considered. When the relative risk possibly varies across studies, one may consider a likelihood ratio test to test the null hypothesis that $\tau_i = 0$. Specifically, let

$$
\hat{LR} = 2 \log \left( \frac{\hat{L}_n(\hat{\theta})}{\hat{L}_n(\tilde{\theta}_{H_0})} \right)
$$

where $\hat{\theta}_{H_0} = \arg\max_{\theta} \hat{L}_n(\theta)$. Here, $\hat{L}_n(\hat{\theta}) = \sum_{i=1}^{n} [Y_i \log \{N_{i1}/(N_{i1} + N_{i2})\} + Y_2 \log \{N_{i2}/(N_{i1} + N_{i2})\}]$ for the conditional approach, and $\hat{L}_n(\tilde{\theta}_{H_0}) = \sum_{i=1}^{n} [z \log(\beta_i) + \log \Gamma(Y_i + z) - \log \Gamma(z) - (Y_i + z) \log(\beta_i + N_{i1} + N_{i2})]$ for the Gamma–Beta approach. Note that under $H_0$, $\psi = \infty$ and $\gamma = 1$. Since one of the parameters under the null is in the boundary of the parameter space, the likelihood ratio statistic $\hat{LR}$ follows a mixture of $\chi^2_2$ distribution with 50:50 of $\chi^2_1$ and $\chi^2_2$.

### 4. Numerical studies

#### 4.1. Example: rosiglitazone study

We applied the various methods described in Section 2 to the rosiglitazone data depicted in Table I. For illustration, we describe our analysis for the MI endpoint only. The table includes the 42 studies reported by Nissen and Wolski [1] as well as 6 additional studies excluded from the Nissen and Wolski analysis because they had no reported events for both MI and CVD related death. The forest plots from the Niessen and Wolski analysis are show in Figure 1. For the MI endpoint, they obtained a 95 per cent confidence interval of $(1.03, 1.98)$ with a $p$-value of 0.03 for the odds ratio between rosiglitazone and the control arm (in favor of the control). The results on the average baseline and relative risks obtained from various methods are summarized in Table II. All of the following confidence intervals are constructed based on normal approximations. For the model with normal prior for $\tau$, the likelihood function was approximated using the Gaussian Quadrature method.

Under the assumption of a constant relative risk as in (2), we estimated the relative risk $e^\tau$ as 1.33 with 95 per cent confidence intervals [0.96, 1.84]. To test whether the relative risk is 1 or equivalently $\tau = 0$, we obtain a $p$-value of 0.087 suggesting that there is marginally significant evidence that the treatment group has a higher risk of MI compared with the control group.

Under the random relative risk framework as in (4) with gamma baseline and log-relative risk following a normal distribution with mean $\mu$ and variance $\sigma^2$, the estimated value of $e^\tau$ was 1.33 with standard error 0.22 and 95 per cent confidence interval [0.96, 1.84]. Similar to the results from the fixed effect model (2), the $p$-value for testing $\tau = 0$ is 0.087. The estimated $\sigma$ was $\hat{\sigma} = 1.31 \times 10^{-6}$ and the standard deviation of the relative risk was $1.26 \times 10^{-6}$. This suggests that there is little heterogeneity in the relative risks across studies.

To allow for more flexible modeling, we also analyzed the data via the FB approach where we assume $\tau_i \sim N(\mu, \sigma^2)$ with $\mu \sim \text{Uniform}(-5, 5)$ and $\sigma \sim \text{Uniform}(0, 2)$ and $\zeta_i \sim \text{Gamma}(x, \beta)$ with $x \sim \text{Uniform}(0.5, 5)$ and $\beta \sim \text{Uniform}(200, 1000)$. These prior distributions for $x$ and $\beta$ were chosen based on the resulting induced prior on the mean and standard deviation of $\zeta_i$ in an attempt to reflective non-informativeness in a rare event setting. The model was fit using WinBUGS with three chains of 30 000 iterations each. Under this approach, the posterior mean for $\mu$ was 0.21 with 95 per cent credible interval $[-0.22, 0.60]$; the posterior mean for $e^{\mu}$ was 1.26 with 95 per cent credible interval [0.80, 1.82]. The $p$-value for testing $\mu$ being 0 is 0.144 suggesting that the elevated MI risk is non-conclusive. The posterior mean for $\sigma$ was 0.28 with 95 per cent credible interval [0.01, 0.79]. The estimated between-study variability from the FB approach is larger compared with that of the above empirical Bayes approach. However, this is likely due to the fact that there is not sufficient information available to estimate the between-study variability well.

Under the conditional approach and assuming that the log-relative risk follows a Beta distribution, maximizing the likelihood function with respect to $\gamma$ and $\psi$, we obtain estimates $\hat{\gamma} = 1.42$ and $\log(\psi) = 17.3$. A 95 per cent confidence
Table I. Data from Nissen and Wolski (2007).

<table>
<thead>
<tr>
<th>Study</th>
<th>Rosiglitazone</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>MI</td>
</tr>
<tr>
<td>1</td>
<td>357</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>391</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>774</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>213</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>232</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>121</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>110</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>382</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>284</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>294</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>563</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>278</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>418</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>395</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>203</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>104</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>212</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>138</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>196</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>122</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>561</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>116</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>148</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>231</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>89</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>168</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>116</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>1172</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>706</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>204</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>288</td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>254</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>314</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>162</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>442</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>394</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>2635</td>
<td>15</td>
</tr>
<tr>
<td>42</td>
<td>1456</td>
<td>27</td>
</tr>
<tr>
<td>43</td>
<td>101</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>232</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>196</td>
<td>0</td>
</tr>
<tr>
<td>48</td>
<td>676</td>
<td>0</td>
</tr>
</tbody>
</table>

Columns show study sizes (N), number of myocardial infarctions (MI) and number of deaths (D) for rosiglitazone and the corresponding control arm for each study.

interval for the location parameter $\gamma$ is $[1.03, 1.95]$. Under the null that the two groups have identical risk, we have $\gamma = 1$. The $p$-value for testing $\gamma = 1$ based on the Wald test is 0.030 suggesting that there is potentially elevated risk of MI in the treated group. In Figure 2, we display the estimated density of $\tau_i W_i$, $f_B(\tau; \hat{\theta})$, for all the observed values of $W_i$ in the rosiglitazone data set. The relative risk was found to be 1.42 and the standard deviation of the relative risk was $5.44 \times 10^{-4}$, which also suggests that there appears to be little heterogeneity across studies. Using the proposed likelihood ratio test for $\tau_0 = 0$, the $p$-value is 0.068. Note that since there is little between-study heterogeneity, such a global test may not be more powerful than the simple Wald test for $\gamma = 1$.  

Figure 1. Ninety-five per cent confidence intervals of the log (odds ratio) for MI and CVD death (control is the reference group) with 48 studies from the Nissen and Wolski analysis (small circles are the observed log (odds ratios). Studies marked with * are those with zero events in both arms for the corresponding outcome.

Table II. Estimates of the mean, median and standard deviation (SD) of the baseline and relative risks based on various methods.

<table>
<thead>
<tr>
<th></th>
<th>Baseline risk×10³</th>
<th>Relative risk</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>SD</td>
<td>Median</td>
</tr>
<tr>
<td>Fixed</td>
<td>2.91</td>
<td>3.75</td>
<td>3.12</td>
<td>1.33</td>
</tr>
<tr>
<td>Normal</td>
<td>2.91</td>
<td>3.75</td>
<td>3.12</td>
<td>1.33</td>
</tr>
<tr>
<td>Random conditional</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.42</td>
</tr>
<tr>
<td>Beta</td>
<td>2.91</td>
<td>3.75</td>
<td>3.12</td>
<td>1.33</td>
</tr>
<tr>
<td>Full Bayes</td>
<td>2.99</td>
<td>3.84</td>
<td>2.27</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Under the gamma–beta framework, the estimated parameters are

\[ \hat{\alpha} = 1.44, \quad \hat{\beta} = 384.8, \quad \log(\hat{\psi}) = 19.8 \quad \text{and} \quad \hat{\gamma} = 1.33. \]

The corresponding 95 per cent confidence interval for \( \gamma \) is [0.96, 1.84] and this approach suggests a marginally significant evidence that there is an elevated risk of MI in the treated group with a \( p \)-value of 0.085 for testing \( \gamma = 1 \). In Figure 3, we display the estimated density of \( \tau_i | W_{ij} = \omega \), \( f_G(\tau | \omega, \theta) \) for all the observed values of \( W_{ij} \) in the rosiglitazone data set. Averaging over all studies, we also obtained the marginal distribution of \( \tau_i \). The estimated median of the marginal distribution was 0.284 with standard error 0.164 and 95 per cent confidence interval [−0.038, 0.61]. The \( p \)-value for testing median log relative risk being 0 is 0.087, which is consistent with the results from the normal prior. The between-study heterogeneity was also very small with a standard deviation of 2.21×10⁻⁴ for the relative risk. Using the proposed likelihood ratio test for \( \tau_i = 0 \), the \( p \)-value is 0.16.

To compare the model fit, we show in Table II the AIC and BIC for all the above four models. Note that the BIC is an approximation to the log marginal likelihood of a model, and therefore, the difference between two BIC estimates may be a good approximation to the natural log of the Bayes factor [20]. Given equal priors for all competing models,
choosing the model with the smallest BIC is equivalent to selecting the model with the maximum posterior probability. The fixed relative risk model with Gamma baseline was preferred over the random relative risks with Gamma baseline. This is expected since little between-study heterogeneity was found in such cases. It is interesting to observe that the conditional model was considered as the best choice for this setting. Our numerical studies (details not reported) suggest that with extremely rare events, we find that the conditional model was often preferred based on these criteria. On the other hand, when the event rate is not too low, the Gamma–Beta approach is often preferred. This is consistent with our expectation of information loss due to the conditioning when there is sufficient data to estimate the baseline event rate well. Although the information criteria selects the conditional approach for the rare event data, results from our simulation studies suggest that the estimator from the conditional approach may be unstable with large bias. Thus caution should be used when interpreting results from the conditional approach.

4.2. Simulation studies

We conducted simulation studies to examine the performance of our proposed procedures in finite samples. Several configurations were considered. The sample sizes are generated by sampling the paired sample sizes from the rosiglitazone data. For a given pair of $N_{i1}$ and $N_{i2}$, we generated the total number of events for the two groups from

$$Y_{i2} \sim \text{Poisson}(N_{i} \lambda_{i2}), \quad Y_{i1} \sim \text{Poisson}(N_{i} \lambda_{i2} e^{\gamma_i}) \quad \text{with} \quad \lambda_{i2} \sim \text{Gamma}(1.44, \beta)$$
with $\beta$ chosen to be either 380 or 38 resulting in a very rare baseline event rate of 0.38 per cent or a moderately rare event rate of 3.8 per cent, respectively. $x=1.44$ and $\beta=380$ were chosen to mimic the rosiglitazone study. Here, the log relative risk $t_i$ was generated based on $t_i = \log(W_i) + \log(\text{Beta}(\gamma, \psi))/2$ for the Beta–Binomial conditional approach and $t_i = \log(W_i) + \log(\text{Beta}(\gamma, \psi)/2)$ for the Gamma–Beta unconditional approach. Generating $t_i$ from these models leads to a Beta prior for $p_i = e^{t_i}/(W_i + e^{t_i})$ and $p_{i\beta} = e^{t_i}/(W_i + e^{t_i})$ for the two approaches. The total number of studies $M$ was chosen to be 48 reflecting the typical number of available studies in practice and 480 reflecting the ideal situation where we expect the asymptotic results to hold. We consider $\gamma=1.33$ (log($\psi$)=5.88) and $\psi=3.58$ (log($\psi$)=1.28), representing a small and large between-study variation, respectively. We let $\gamma=1.33$ which correspond to median relative risk of about 1.33 for both the Beta–Binomial model and the Gamma–Beta model. The interquartile range for the relative risk is around 0.10 when log($\psi$)=5.88 and 1.10 when log($\psi$)=1.28. The empirical biases, the sampling standard errors, and the average estimated standard errors for $\gamma$ and log($\psi$) are summarized in Table III. The results are based on 2000 simulated data sets. Note that the proportion studies with zero events in both arms are about 4 per cent (about 2 out 48) when $\beta=38$ and 34 per cent (about 16 out of 48) when $\beta=380$.

First, consider the setting when the baseline event has a very low rate of 0.38 per cent. In the absence of substantial between-study variation, both methods estimate the center parameter $\gamma$ well with negligible bias and almost identical standard errors. Interestingly, this is consistent with the results from the fixed effects framework. The conditional and unconditional approaches result in almost identical estimates of the relative risks. On the other hand, both methods appear to have difficulty in estimating log($\psi$). This is not surprising since $1/\psi$, reflecting the between-study variation, is close to the boundary of the parameter space. When we increase the between-study variation and log($\psi$)=1.28, the unconditional Gamma–Beta approach clearly becomes preferable to the conditional Beta–Binomial approach. The conditional approach tends to produce biased estimates even with 480 studies. For the estimation of $\gamma$ with $M=48$, the conditional approach results in a bias of 0.49 and mean squared error (MSE) of 0.47, whereas the unconditional approach results in a bias of 0.08 and MSE of 0.14. Thus, the unconditional approach is more than three times as efficient as the conditional approach with respect to MSE. For the estimation of log($\psi$) with $M=48$, the conditional approach appears to be more stable compared with the unconditional approach with much smaller bias. On the other hand, with $M=480$, both approaches provide stable estimates with the unconditional approach being preferable with a smaller MSE of 0.06 compared with the MSE of 0.28 with the conditional approach.

When the baseline event has a moderately low rate of 3.8 per cent, we observe similar patterns as described above for the case with small between-study variation. Both approaches seem to be able to estimate $\gamma$ and log($\psi$) better compared with the rare event case. Both approaches produce stable and similar estimates for $\gamma$. The estimates for log($\psi$) remain unstable when there is little between-study heterogeneity. When the between-study variation is larger, both approaches could estimate $\gamma$ and log($\psi$) well. The conditional approach continues to produce slight biases in the estimated $\gamma$. The unconditional approach yields stable estimators of $\gamma$ and log($\psi$) with negligible biases. When $M=480$, the MSEs for (log($\gamma$), log($\psi$)) are (0.031, 0.020) for the conditional approach and (0.008, 0.004) for the unconditional approach. Thus, with respect to MSE, the unconditional approach is still significantly more efficient compared with the conditional approach.

Across all settings considered, the estimated standard errors are close to the sampling standard errors for $\gamma$ suggesting the validity of using observed information matrix for making inference in finite sample. When there is significant between-study variation, the standard error estimates for log($\psi$) are also close to observed sampling standard errors. However, inference for log($\psi$) appears to be unreliable when the true log($\psi$) is large, i.e. when there is little between-study variation.

<table>
<thead>
<tr>
<th>$\beta=380$</th>
<th>$\beta=38$</th>
<th>$\beta=380$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M=48$</td>
<td>$M=48$</td>
<td>$M=480$</td>
</tr>
<tr>
<td>$M=480$</td>
<td>$M=480$</td>
<td>$M=480$</td>
</tr>
<tr>
<td>$\gamma=1.33$</td>
<td>$\gamma=1.33$</td>
<td>$\gamma=1.33$</td>
</tr>
<tr>
<td>$\log(\psi)=5.88$</td>
<td>$\log(\psi)=5.88$</td>
<td>$\log(\psi)=1.28$</td>
</tr>
<tr>
<td>Bias</td>
<td>SSE</td>
<td>ASE</td>
</tr>
<tr>
<td>0.03</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>0.02</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>0.01</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>0.00</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>0.01</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>0.00</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The estimates for $\gamma$ and log($\psi$) are well with negligible bias and almost identical standard errors. The conditional and unconditional approaches result in almost identical estimates of the relative risks. On the other hand, both methods appear to have difficulty in estimating log($\psi$). This is not surprising since $1/\psi$, reflecting the between-study variation, is close to the boundary of the parameter space. When we increase the between-study variation and log($\psi$)=1.28, the unconditional Gamma–Beta approach clearly becomes preferable to the conditional Beta–Binomial approach. The conditional approach tends to produce biased estimates even with 480 studies. For the estimation of $\gamma$ with $M=48$, the conditional approach results in a bias of 0.49 and mean squared error (MSE) of 0.47, whereas the unconditional approach results in a bias of 0.08 and MSE of 0.14. Thus, the unconditional approach is more than three times as efficient as the conditional approach with respect to MSE. For the estimation of log($\psi$) with $M=48$, the conditional approach appears to be more stable compared with the unconditional approach with much smaller bias. On the other hand, with $M=480$, both approaches provide stable estimates with the unconditional approach being preferable with a smaller MSE of 0.06 compared with the MSE of 0.28 with the conditional approach.

When the baseline event has a moderately low rate of 3.8 per cent, we observe similar patterns as described above for the case with small between-study variation. Both approaches seem to be able to estimate $\gamma$ and log($\psi$) better compared with the rare event case. Both approaches produce stable and similar estimates for $\gamma$. The estimates for log($\psi$) remain unstable when there is little between-study heterogeneity. When the between-study variation is larger, both approaches could estimate $\gamma$ and log($\psi$) well. The conditional approach continues to produce slight biases in the estimated $\gamma$. The unconditional approach yields stable estimators of $\gamma$ and log($\psi$) with negligible biases. When $M=480$, the MSEs for (log($\gamma$), log($\psi$)) are (0.031, 0.020) for the conditional approach and (0.008, 0.004) for the unconditional approach. Thus, with respect to MSE, the unconditional approach is still significantly more efficient compared with the conditional approach.

Across all settings considered, the estimated standard errors are close to the sampling standard errors for $\gamma$ suggesting the validity of using observed information matrix for making inference in finite sample. When there is significant between-study variation, the standard error estimates for log($\psi$) are also close to observed sampling standard errors. However, inference for log($\psi$) appears to be unreliable when the true log($\psi$) is large, i.e. when there is little between-study variation.
Table IV. Empirical Size/Power of the six tests, likelihood ratio for Beta–Binomial (LRBB), likelihood ratio for Gamma–Beta (LRGB), fixed effects Mantel Hansel with 0 imputation (MH\(_{0}^{F}\)) and with 0.5 imputation (MH\(_{0.5}^{F}\)), DerSimonian–Laird random effects with 0.5 imputation (DL\(_{0.5}^{R}\)), Peto’s method with fixed (PF) and with random effects (PR), fully Bayesian approach (FB), as well as the fixed (Fix) relative risk approach with Gamma baseline for three settings: (a) \(e_{i}^{r} = 1\); (b) \(e_{i}^{r} \sim\) Gamma(66.5, 50); and (c) \(e_{i}^{r} \sim\) Gamma(0.665, 0.5).

<table>
<thead>
<tr>
<th></th>
<th>LRBB</th>
<th>LRGB</th>
<th>MH(_{0}^{F})</th>
<th>MH(_{0.5}^{F})</th>
<th>DL(_{0.5}^{R})</th>
<th>PF</th>
<th>PR</th>
<th>FB</th>
<th>Fix</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.44</td>
<td>0.40</td>
<td>0.054</td>
<td>0.046</td>
<td>0.034</td>
<td>0.056</td>
<td>0.043</td>
<td>0.042</td>
<td>0.055</td>
</tr>
<tr>
<td>M = 480</td>
<td>0.045</td>
<td>0.048</td>
<td>0.051</td>
<td>0.142</td>
<td>0.192</td>
<td>0.051</td>
<td>0.047</td>
<td>0.078</td>
<td>0.052</td>
</tr>
<tr>
<td>b</td>
<td>0.254</td>
<td>0.269</td>
<td>0.320</td>
<td>0.169</td>
<td>0.093</td>
<td>0.324</td>
<td>0.288</td>
<td>0.109</td>
<td>0.338</td>
</tr>
<tr>
<td>M = 480</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
<td>0.949</td>
<td>0.785</td>
<td>0.998</td>
<td>0.998</td>
<td>0.974</td>
<td>0.998</td>
</tr>
<tr>
<td>c</td>
<td>0.833</td>
<td>0.914</td>
<td>0.422</td>
<td>0.355</td>
<td>0.102</td>
<td>0.422</td>
<td>0.101</td>
<td>0.233</td>
<td>0.430</td>
</tr>
<tr>
<td>M = 480</td>
<td>1.00</td>
<td>1.00</td>
<td>0.875</td>
<td>0.742</td>
<td>0.122</td>
<td>0.875</td>
<td>0.076</td>
<td>0.932</td>
<td>0.880</td>
</tr>
</tbody>
</table>

To test the null hypothesis that there is no treatment effect, we compared our proposed likelihood ratio test to the fixed effects approach with Gamma baseline as well as to the standard approaches based on the fixed effects MH with and without 0.5 imputation for zero cells, the DerSimonian and Laird (DL) random effects which use 0.5 imputation as well as Peto’s method with fixed effects (PF) and with random effects (PR). For the PR, Peto’s log odds ratio and its standard error are utilized in the random effects model. The detailed implementation of such random effects approaches can be found in [21]. For comparison, we also performed testing based on the FB approach with prior distributions selected the same as described above. To mimic the diabetes example, we generated the data with sample sizes \(N_{11}, N_{12}\) fixed the same as in the example. For each study, we generated \(Y_{12}\) from Poisson(\(N_{11} N_{12}\)) and \(Y_{11}\) from Poisson(\(N_{11} N_{12} e^{2}i\)) with \(\lambda_{2} \sim\) Gamma(1.44, 380) and \(e_{i}\) generated from various Gamma distributions to reflect different degrees of treatment effect. For each of the configuration, we generated 2000 data sets and summarized the empirical size or power of all the tests at the type I error rate of 0.05.

We first assessed the validity of the tests in finite sample by generating the data under the null setting (a) with \(\tau_{i} = 0\) or \(e_{i}^{r} = 1\). As shown in Table IV, the empirical sizes of the proposed likelihood ratio tests are reasonably close to the nominal level. The conditional approach yields a size of 0.044 and the Gamma–Beta approach yields a size of 0.040 when \(n = 48\). Similarly, sizes are observed for the standard methods except that both the MH and DL methods with artificial 0.5 imputations yielded inflated sizes when \(M = 480\). This raises the concern of applying arbitrary imputations, which has been noted in existing studies [5, 6]. The empirical size of the FB approach also appears to be slightly higher than the nominal level.

To compare the power of the tests, we considered two different settings: (b) \(e_{i}^{r} \sim\) Gamma(66.5, 50); and (c) \(e_{i}^{r} \sim\) Gamma(0.665, 0.5). Both settings yield an average relative risk of 1.33, but the between-study variability is small (std deviation = 0.16) in setting (b) and is large (std deviation = 1.63) in setting (c). With a small between-study variability, testing under the fixed effects framework may be more powerful since additional parameter estimations are required in the random effects framework. This is consistent with the results shown in Table IV. All the fixed effects procedures without imputations are slightly more powerful compared with the proposed likelihood ratio tests. The MH method with 0.5 imputation appears to have much less power compared with other methods. This is likely due to the imputation of 0.5 which may result in a significant variance over-estimation for rare event settings. Furthermore, when \(n = 48\), all the tests have rather limited power (ranging from 9.3 per cent to 33.8 per cent) in detecting the treatment difference. With a large between-study variation in setting (c), one may fail to detect the treatment difference by only assessing the center of the relative risks. With \(n = 48\), the fixed effects approach yield tests with 42 per cent of power for the Peto’s method, 36 per cent of power for MH method, and 43 per cent of power for the fixed effects approach with Gamma baseline. On the contrary, the proposed likelihood ratio tests yield 83 per cent of power based on the conditional approach and 91 per cent of power based on the Gamma–Beta approach. Furthermore, the standard approaches with random effects including the FB approach have only about 10 per cent of power in detecting the difference when the tests are only focused on the mean relative risks. This is partially due to the fact that the standard approaches focus on the testing of \(\mu = 0\) as for fixed effects models without restricting the null to be \(\tau_{i} = 0\) across all studies.

5. DISCUSSION

In this paper, we proposed the use of Poisson random effects model to combine \(2 \times 2\) tables from multiple studies when the event rates for individual studies are rare. In the presence of between-study heterogeneity, the mean log-relative risk can only be viewed as the center of the distribution of the treatment effect. For such settings, it is important to
assess the amount of between-study heterogeneity in the log-relative risk which also captures information regarding the treatment effect. The proposed likelihood ratio tests are useful tools for assessing the overall treatment effects. Simulation results suggest that the proposed tests are more powerful compared with the standard tests when there is a high level of between-study variation. We took advantage of the conjugacy of the Poisson and Gamma distributions, as well as of Beta and the Binomial distributions to obtain closed-form likelihood functions. One potential drawback of the proposed approach is that the Beta prior imposed on $\pi_i$ and $p_{ji}$ implicitly specifies the prior of $\tau_i$ to be dependent on the relative sample size allocation. In practice, the maximum likelihood estimators may be obtained easily through standard optimization algorithms such as nlmin in R or the Newton Raphson algorithm. The R code for implementing the proposed procedures is available upon request.

The proposed procedures can be easily extended to incorporate additional covariates. The study level covariates such as treatment duration could be incorporated easily. Covariates thought to influence baseline event rates could be simply added as linear terms into the basic log-linear model (1) on $\lambda_{ij}$. Covariates thought to influence that the relative risk itself could be incorporated as interaction effects. The methods could also be extended to analyze individual patient level data to incorporate patient level covariates such as age or gender. For such an extension, one may consider the Poisson model as an approximation to the binomial model with rare events and model the covariate effect multiplicatively on the probability of experiencing an adverse event. Although a large population of studies would be needed in order to have precision to assess the impact of covariates, such a meta-regression approach would, in theory, provide a highly valuable strategy for exploring whether the treatment differences vary across sub-populations. Such explorations could be used to identify sub-populations with particularly high or low risks of adverse events. The proposed procedures could also be extended to analyze time to event data by extending the Poisson model to the exponential model. Consider the simple setting without additional covariates, the likelihood contribution from the $j$th treatment group of the $i$th study is simply $\lambda_{ij}^{d_{ij}}e^{-\lambda_{ij}}$, where $\lambda_{ij} = \sum_{k=1}^{N_j} \delta_{ijk}$ and $\bar{\lambda}_{ij} = \sum_{k=1}^{N_j} (1 - \delta_{ijk}) \lambda_{ijk}$. $\delta_{ijk}$ is the indicator of experiencing an event during the follow up time and $\lambda_{ijk}$ is the minimum of the event time and the follow up time for the $k$th subject in the study. Our procedure would follow immediately by replacing the Poisson likelihood with the above likelihood.

Through simulation studies, we find that the conditional approach, which essentially deletes studies with zero events in both treatment arms, may result in biased estimates when the underlying event rate is low and the between-study variation is not small. However, the bias does diminish as we increase the study sample sizes and the type I error for testing $\gamma = 1$ is close to the nominal level when the study sample sizes are increased to an average level of 2000. On the other hand, the unconditional approach, which utilizes information across all studies, performs well in estimating the relative risk across all the settings considered in our numerical studies. In general, we find that the unconditional approach is more efficient than the conditional approach when there is a high level of between-study variation and the two approaches have similar performances when there is little between-study variation. It is worth noting that in order to estimate the between-study variation well, one may require the study sample sizes and the number of studies to be large, especially if the baseline event rate is extremely low.

Acknowledgements

This work was supported in part by grants CA48061 and I2B2 from the National Institutes of Health. The authors are grateful to Professor L.J. Wei for his inspiring discussions and helpful suggestions.

References