Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 259–283
http://www.ssca.org.in/journal

Meta Analysis for Rare Events

Dulal K. Bhaumik[1,2], **Anup K. Amatya**[1,2] **and Soumya Sahu**[1]
[1]*Division of Epidemiology and Biostatistics, University of Illinois Chicago*
[2]*Department of Psychiatry, University of Illinois Chicago*

Received: 30 April 2024; Revised: 09 July 2024; Accepted: 28 August 2024

Abstract

Meta-analysis has become a widely used tool for evaluating the efficacy and safety of medical interventions, offering numerous advantages and utilities. However, recent studies have raised questions about the accuracy of commonly used moment-based meta-analytic methods, particularly for rare binary outcomes. This issue is further complicated in studies with heterogeneous effect sizes. Likelihood-based mixed-effects modeling provides an alternative to moment-based methods, such as inverse-variance weighted fixed- and random-effects estimators. In this review paper, we discuss several meta-analysis methods specifically designed for analyzing rare event data. We elaborate on the use of continuity correction for studies with zero total events, taking into account study heterogeneity. The problem is motivated, and results are illustrated using a well-known meta-analysis study. By exploring and comparing these different methodologies, researchers can gain insights into the most appropriate approaches for analyzing rare event data in meta-analytic studies.

Key words: Conditional likelihood; Mantel-Haenszel method; The Peto method; Confidence distribution methods; Odds ratio.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

Meta-analysis is a powerful statistical tool used to combine results from multiple studies, particularly useful for making robust inferences about rare events, which require large sample sizes due to their low frequency (less than 0.1%). Traditional clinical trials often lack sufficient power to draw sound conclusions about rare adverse events, such as those associated with pharmaceutical agents. The challenge lies in incorporating studies with few or no observed adverse events into the analysis. While fixed-effect and random-effect meta-analyses are common, Bayesian methodologies and confidence distribution approaches offer alternatives. Each method has unique strengths and weaknesses, and the optimal approach for analyzing rare events remains a topic of ongoing research. We try to clarify the idea that rare event meta analysis may end up with some studies with zero total events. However, those studies with zero events are also informative and should be included in the analysis

Corresponding Author: Dulal K. Bhaumik
Email: dbhaumik@uic.edu

and this demands the concept of continuity correction. Data analysis of meta analysis of rare events is developed addressing these concerns and difficulty of zero total events depending on what type methodology is used.

Meta-analysis is a convenient statistical tool that combines results from multiple trials and makes a robust inference regarding the parameter of our interest. To make a valid statistical inference for a rare event requires a trial with a large sample size. Regular clinical trials are not sufficiently powered to draw a statistically sound conclusion regarding events that often occur at a rate of less than 0.1%. Low-frequency events are commonly encountered in the investigation of adverse events (*e.g.*, suicide) associated with a pharmaceutical agent (*e.g.*, antidepressants). A further complication arises in rare adverse event studies because clinical trials are typically designed to assess the efficacy rather than the safety of the product. As a result, the chance of observing a reasonable number of such adverse events in a single trial or study is relatively low. Quite often, in such situations, not even a single adverse event is observed in efficacy trials. Utilizing such studies meaningfully in the analysis is the greatest challenge in the meta-analysis of rare events. Several strategies have been proposed to make a valid decision regarding the parameters of our interest, incorporating all available studies. However, none of those is universally accepted, and as a result, the issue is still open. Traditional methods of meta-analysis either treat the underlying treatment effect as a fixed parameter across multiple studies or assume individual study treatment effect as a random sample from a hypothetical pool of treatment effects. The first form of meta-analysis is called fixed-effect meta-analysis, and the second form is called random-effect meta-analysis. Bayesian methodologies are also used to allow hierarchical modeling with a greater opportunity for sensitivity analysis. Recently, the third method, based on the concept of confidence distribution, has been put forth as an attractive alternative for meta-analysis of rare events. For each methodology, there are several estimation techniques with respective strengths and weaknesses. This article discusses some characteristics of rare event studies and provides an overview of meta-analytic methods suitable for the analysis of rare events, along with the issues pertaining to those methods.

2. Zero total event studies

A study in which no outcome event is observed is called a zero-total event study. Studies where outcome events are observed in one arm but not in the other arm are called zero-cell studies. Zero total event studies in rare event analysis are contentious due to their lack of events in one or both treatment arms, but recent literature suggests they should not be excluded, despite challenges in variance computation and continuity correction. The ubiquitous characteristic of rare event studies is the absence of events in either one or both treatment arms. The answers have been contentious, and inconclusive. The core of the issue is the argument that the zero total event studies do not contribute any information towards the estimation of the effect and, hence, are irrelevant and should be removed from the analysis Whitehead and Whitehead (1991); Sweeting *et al.* (2004). However, in general, a zero total event study with a large sample size is expected to provide stronger evidence for any hypothesized effect compared to a smaller sample size zero total study Friedrich *et al.* (2007); Liu (2012); Kuss (2014). Furthermore, recent publications are providing theoretical support to the relevance of the zero total event studies Liu *et al.* (2014); Xie *et al.* (2014). Therefore, zero total event studies should not be excluded. The major obstacle to the inclusion of

zero total event studies in a traditional meta-analysis is the numerical ill-conditioning for the computation of variance of the effect size using ratio measures. Analysts are addressing this problem by proposing the concept of Continuity Correction, even though there is no consensus on what exact value we should use for the continuity correction. Additional complexity arises when a significant heterogeneity exists among studies.

Example: Zero total event studies in meta-analysis of rositaglitazone and risk of cardiovascular events

On November 14, the Food and Drug Administration (FDA) put a "black box warning" on Rositaglitazone's product to inform consumers of such risks. On September 23, 2010, the FDA limited access to Rositaglitazone because of concerns about increased cardiovascular risk. The most prominent study that led to the action by the FDA was the meta-analysis conducted by Nissen and Wolski (2007). As part of the analysis, 42 trials were selected from the published literature, the FDA website, and a clinical trial registry maintained by the drug manufacturer (GlaxoSmithKline). Table 1 reports the myocardial infarction (MI) events and deaths from cardiovascular causes (CVD) that were reported in the 42 clinical trials included in the study. Of those 42 studies, four studies (9.5%; study 20, 31, 33, and 38) are zero total event studies for MI endpoint, and 19 studies (45%; study 2–4, 6, 7, 9, 10, 12, 14, 17, 21–24, 29, 31, 36–38) are zero total event studies for CVD endpoint. Overall, there were 86 MIs and 39 CVDs in the rosiglitazone group and 72 MIs and 39 CVDs in the control group.

2.1. Conditional likelihood

This section explains the practical reasons for not favoring the zero total event studies, mainly because of computational difficulty under the set up of conditional likelihood. The most compelling argument for supporting the exclusion of zero total event studies comes from the conditional likelihood inference perspective. The conditional maximum likelihood estimation procedure estimates the parameter of interest by maximizing conditional likelihood given the minimal sufficient statistics for the nuisance parameters. Consider a sequence of observations $\{x_{t1}, x_{t2}, \ldots, x_{tk}\}$, and $\{x_{c1}, x_{c2}, \ldots, x_{ck}\}$ from k studies/trials with $\{n_{t1}, n_{t2}, \ldots, n_{tk}\}$, and $\{n_{c1}, n_{c2}, \ldots, n_{ck}\}$ treatment, and control group sample sizes respectively. Observations from an individual study form the following 2×2 table given in Table 2.

For a fixed observed event total t_i, only random variable in the ith table is X_i ($i = 1, 2, \ldots, k$) (count in the upper left cell), which follows a hyper-geometric distribution. The corresponding conditional likelihood function given $T_i = t_i$ is as follows:

$$L_{x_{ti}|t_i}(\theta) = P_\theta(X_{ti} = x_{ti}|T_i = t_i) = \frac{\binom{n_{ti}}{x_{ti}}\binom{n_{ci}}{t_i - x_{ti}}\psi^{x_{ti}}}{\sum_{\nu=u_i}^{l_i}\binom{n_{ti}}{\nu}\binom{n_{ci}}{t_i-\nu}\psi^{\nu}}, \tag{1}$$

and the joint conditional likelihood function is given by the following expression:

$$\phi(x_{t1}, x_{t2}, \ldots, x_{tk}|t_1, t_2, \ldots, t_k) = \prod L_{x_{ti}|t_i}(\theta), \tag{2}$$

Table 1: Example data: Rositaglitazone and the risk of cardiovascular events Nissen and Wolski (2007)

study	Rosiglitazone			Control		
	Total	MI	CVD	Total	MI	CVD
1	357	2	1	176	0	0
2	391	2	0	207	1	0
3	774	1	0	185	1	0
4	213	0	0	109	1	0
5	232	1	1	116	0	0
6	43	0	0	47	1	0
7	121	1	0	124	0	0
8	110	5	3	114	2	2
9	382	1	0	384	0	0
10	284	1	0	135	0	0
11	294	0	2	302	1	1
12	563	2	0	142	0	0
13	278	2	0	279	1	1
14	418	2	0	212	0	0
15	395	2	2	198	1	0
16	203	1	1	106	1	1
17	104	1	0	99	2	0
18	212	2	1	107	0	0
19	138	3	1	139	1	0
20	196	0	1	96	0	0
21	122	0	0	120	1	0
22	175	0	0	173	1	0
23	56	1	0	58	0	0
24	39	1	0	38	0	0
25	561	0	1	276	2	0
26	116	2	2	111	3	1
27	148	1	2	143	0	0
28	231	1	1	242	0	0
29	89	1	0	88	0	0
30	168	1	1	172	0	0
31	116	0	0	61	0	0
32	1172	1	1	377	0	0
33	706	0	1	325	0	0
34	204	1	0	185	2	1
35	288	1	1	280	0	0
36	254	1	0	272	0	0
37	314	1	0	154	0	0
38	162	0	0	160	0	0
39	442	1	1	112	0	0
40	394	1	1	124	0	0
41	2635	15	12	2634	9	10
42	1456	27	2	2895	41	5

Table 2: 2×2 **contingency table for the** ith **trial/study**

	Event		
	yes	no	total
Treatment	x_{ti}	$n_{ti} - x_{ti}$	n_{ti}
Control	x_{ci}	$n_{ci} - x_{ci}$	n_{ci}
Total	t_i	$(n_{ti} + n_{ci}) - t_i$	$n_{ti} + n_{ci}$

for $l_i \leq x_{ti} \leq u_i$, where $l_i = max(0, t_i - n_{ci})$, $u_i = min(n_{ti}, t_i)$, and $\psi = \exp\theta$ is the odds ratio which is assumed to be the same across k studies involved in the analysis. The value of θ that maximizes this conditional likelihood is called the conditional maximum likelihood estimate (CMLE) for ψ. Note that $L_i\psi(x_{ti}|t_i) = 1$, for zero total event studies, does not directly contribute to the joint conditional likelihood. An asymptotic property of CLME can be proved under a reasonable set of conditions. However, unlike the direct maximum likelihood estimate, the CMLE, in general, is not efficient except for some special (but important) situations, where the asymptotic variance attends the Cramer-Rao lower bound (Andersen, 1970, see). Unfortunately, the CLME obtained from (2) is not derived from one of those special situations and is not an efficient estimator of ψ. Thus, the most reasonable basis for the exclusion of zero total event studies is based on a procedure that maintains asymptotic properties and does not use all information contained in the data for the parameter of interest. Furthermore, Xie *et al.* (2014) has shown conclusively that the zero total event studies do contain information that is a function of ψ, π_{ci}, and sample sizes n_i.

The basic idea behind the derivation outlined by Xie *et al.* (2014) is as follows. Suppose that X_{ti} and X_{ci} are independent binomial random variables following $B(\pi_{ti}, n_{ti})$, and $B(\pi_{ci}, n_{ci})$, respectively. The full (unconditional) likelihood function is given as:

$$L_{\boldsymbol{x_t},\boldsymbol{x_c}}(\theta, \boldsymbol{\pi_c}) = L_{\boldsymbol{x_t},\boldsymbol{x_c}}(\boldsymbol{\pi_t}, \boldsymbol{\pi_c}) = \prod_{i=1}^{k} \binom{n_{ti}}{x_{ti}} \binom{n_{ci}}{x_{ci}} \pi_{ti}^{x_{ti}} (1-\pi_{ti})^{n_{ti}-x_{ti}} \pi_{ci}^{x_{ci}} (1-\pi_{ci})^{n_{ci}-x_{ci}}. \quad (3)$$

Under the assumption that the odds ratio is the same across k studies, π_{ti} and π_{ci} satisfy a constraint $\{\pi_{ti}/(1-\pi_{ti})\}/\{\pi_{ci}/(1-\pi_{ci})\} = e^{\theta}$. From the likelihood principle, it follows that the above likelihood function contains all information relevant for making an inference for the parameter of interest. The full likelihood (3) can be rewritten as

$$L_{\boldsymbol{x_t},\boldsymbol{x_c}}(\theta, \boldsymbol{\pi_c}) = L_{\boldsymbol{x_t}|\boldsymbol{t}}(\theta) D_t(\theta, \boldsymbol{\pi_c}), \quad (4)$$

where $D_t(\theta, \boldsymbol{\pi_c}) = \dfrac{L(\theta, \boldsymbol{\pi_c})}{L_{x|t}(\theta)}$. Xie *et al.* (2014) showed that $D_t(\theta, \boldsymbol{\pi_c})$ is a function of both θ and $\boldsymbol{\pi_c}$. Therefore, they argued that the conditional likelihood inference and full likelihood inference are different, suggesting that "the conditional likelihood approach can incur omission or distortion of information". Clearly, the zero total event studies contribute $\prod_{\{i:t_i=0\}} (1-\pi_{ti})^{n_{ti}} (1-\pi_{ci})^{n_{ci}}$ portion of information to the full likelihood. But that portion of information, which is also a function of both θ and $\boldsymbol{\pi_c}$, is omitted from conditional likelihood. As a result, inferences under conditional likelihood that effectively omit zero total event studies will be weaker and less reliable.

The conditional likelihood (1) is developed under a specific assumption that t_i's are fixed in addition to the same assumption on n_{ti} and n_{ci} for each study. However, in general, studies or trials that are included in meta-analysis do not have control over observed total events. Consequently, the hypothesis testing under the assumption of fixed t_i is conservative and loses power when only the n_{ti} and n_{ci} are fixed. Thus, Xie *et al.* (2014) concluded that zero total event studies do contain information on the intervention effect.

As described above, arguments for and against of excluding zero total event studies are generally put forth by assuming a common odds ratio across studies. However, in contrast to the conservative findings under such assumptions, simulation studies have suggested that methods that exclude zero total event studies can have an inflated type I error rate when odds ratios vary between studies Bhaumik *et al.* (2012). Furthermore, popular methods used in practice have a tendency to overestimate the true odds ratio and underestimate the between study heterogeneity. This also indicates that the zero total event studies do contain relevant information on the parameters of our interest. In what follows, we discuss how to include the zero total event studies in a meaningful way in meta-analysis.

3. Moment matching methods

In this section we discuss some frequently used meta analysis methods based on weighted average estimates with the continuity correction when applied in sparse data. Traditional meta-analysis methods are perhaps the most useful methods that are used in practice. Those are derived based on the moment-matching approach. These methods include various forms of weighted average estimates of the overall intervention effect. The inverse variance weighted method, Mantel-Haenszel method, and Peto method are the three most widely used methods under this category. These methods typically require some form of adjustment when applied to sparse data. Although intended for different purposes in the context of a chi-square test, such adjustment made in individual cells of 2×2 tables in meta-analysis is known as the continuity correction.

3.1. Continuity correction

The controversy over continuity correction in meta-analysis of rare event studies persists, with alternative correction factors proposed to mitigate bias and coverage issues, while recent developments suggest methods avoiding continuity correction altogether could be possible. As mentioned before, continuity correction is a controversial topic. There are competing views on the appropriateness of the use of continuity correction in meta-analysis. In the context of traditional analysis, there is no other choice but to discard zero total event studies or to use a Bayesian approach without any continuity correction. The value that has received the most attention for the continuity correction is 1/2. It was accepted as the value for continuity correction on the basis of the argument put forth in Cox (1970). According to Cox, when using the odds as the effect measure, choosing a correction factor of 1/2 gives the least biased estimator of the true log odds in a single treatment group situation. The factor 1/2 is also used to improve the approximation of a discrete distribution by a continuous distribution (i.e. 1-degrees of freedom chi-square), or to obtain an approximation to the product hypergeometric probability. However, adding a constant continuity correction such as 1/2 can create some undesirable problems, including reversal of the effect direction, particularly if the treatment arms are unbalanced Rücker *et al.* (2009). An investigation by

Sweeting *et al.* (2004) concluded that using the continuity correction of 1/2 may be outperformed in terms of both bias and coverage by other choices of correction factor when studying the odds ratio between two groups. An important point to be noted here is that the study was conducted under the assumption of fixed intervention effect across studies, and excluding zero total event studies. They noted that the application of their alternative continuity correction factor might not be applicable when using random-effect models. Two alternative correction strategies that Sweeting *et al.* (2004) showed to be outperforming, under fixed-effect assumption, are (1) to add a factor of the reciprocal of the size of the opposite treatment arm to the cells, and (2) to use empirical continuity correction. However, they also cautioned that not a single correction factor or method is superior in all situations, and recommended to perform sensitivity analysis using several different correction factors. See Sweeting *et al.* (2004) for details on the aforementioned alternative approaches for continuity correction.

Recent efforts on methodological development and validation studies suggest that the issue of continuity correction can potentially be avoided altogether using those methods (in the frequentist domain) that do not require continuity correction. Furthermore, these methods allow the inclusion of all studies in meta-analysis. Nonetheless, the Mantel-Hanszel and Peto methods are widely used for meta-analysis of rare events. Therefore, these popular classical methods, along with a somewhat underutilized but highly relevant method using arcsine risk difference measure, are briefly described in the following sections.

3.2. Mantel-Haenszel method

The Mantel-Haenszel method for meta-analysis adjusts for potential confounding factors and uses weighted averages to estimate the combined odds ratio, with alternative continuity corrections recommended to mitigate bias and improve coverage. The Mantel-Haenszel method was originally developed for stratified analysis adjusting for the third potential confounding factor. The fixed-effect meta-analysis can be viewed as a stratified design where each individual study is treated as a stratum. Based on the Mantel-Haenazel method, the pooled odds ratio across all K studies is estimated using the following expression:

$$\widehat{OR}_{MH} = \frac{\sum_{i=1}^{K} x_{Ti}(n_{Ci} - x_{Ci})/N_i}{\sum_{i=1}^{K} x_{Ci}(n_{Ti} - x_{Ti})/N_i}. \tag{5}$$

Equation (5) can be rewritten as a weighted average estimate as follows:

$$\widehat{OR}_{MH} = \frac{\sum_{i=1}^{K} w_i \widehat{OR}_i}{\sum_{i=1}^{K} w_i}, \tag{6}$$

$$\text{where } \; w_i = \frac{x_{Ci}(n_{Ti} - x_{Ti})}{N_i}, \text{ and } \widehat{OR}_i = \frac{x_{Ti}(n_{Ci} - x_{Ci})}{x_{Ci}}(n_{Ti} - x_{Ti}) \tag{7}$$

It is clear from equation (5) that zero cell studies contribute to the estimation of a combined odds ratio. However, zero total event studies are implicitly excluded from the computation unless a continuity correction is added. The weights in equation (7) are not reciprocals of the variances of odds ratio estimates from individual studies. Therefore, the variance estimate

of the combined odds ratio is not as straightforward as in the inverse variance method. The Robins-Breslow-Greenland method is generally accepted as an easy-to-use variance estimator for $ln(\widehat{OR}_{MH})$. It has the following expression:

$$\widehat{Var}[ln(\widehat{OR}_{MH})] = \frac{S_3}{2S_1^2} + \frac{S_5}{2S_1S_2} + \frac{S_4}{2S_2^2}, \tag{8}$$

where $S_1 = \sum_{i=1}^{K} \frac{x_{Ti}(n_{Ci} - x_{Ci})}{N_i}$, $S_2 = \sum_{i=1}^{K} \frac{x_{Ci}(n_{Ti} - x_{Ti})}{N_i}$,
$S_3 = \sum_{i=1}^{K} \frac{x_{Ti}(n_{Ci} - x_{Ci})(x_{Ti} + n_{Ci} - x_{Ci})}{N_i^2}$, $S_4 = \sum_{i=1}^{K} \frac{x_{Ci}(n_{Ti} - x_{Ti})(x_{Ci} + n_{Ti} - x_{Ti})}{N_i^2}$,
and $S_5 = \sum_{i=1}^{K} \frac{x_{Ci}(n_{Ti} - x_{Ti})(x_{Ti} + n_{Ci} - x_{Ci}) + x_{Ti}(n_{Ci} - x_{Ci})(x_{Ci} + n_{Ti} - x_{Ti})}{N_i^2}$. A null hypothesis of equal odds in treatment and control subjects, i.e., $OR_{MH} = 1$, may be tested by the following χ^2-test:

$$X^2_{MH} = \left[\sum_{i=1}^{K} \frac{x_{Ti}(n_{Ci} - x_{Ci}) - x_{Ci}(n_{Ti} - x_{Ti})}{N_i}\right]^2. \tag{9}$$

The Mantel-Haenszel method with the continuity correction of 1/2 produces biased estimates and low coverage rates for event rates below 1 percent Bradburn *et al.* (2007). Therefore, under fixed-effect conditions, an alternative continuity correction is recommended instead of 1/2 to reduce bias and improve coverage characteristics of this estimator Sweeting *et al.* (2004).

3.3. The Peto method

The Peto method in meta-analysis of moderately rare events excludes zero total event studies automatically and estimates the pooled log odds ratio based on weighted differences from individual tables, with limitations in unbalanced data and close-to-1 odds ratios. The Peto method is popular for meta-analysis of moderately rare events. Similar to the Mantel-Haenszel method, this method does not require artificial continuity correction when events are not observed in one of the treatment arms. However, the zero total event studies are automatically given zero weight and effectively are excluded from the analysis. When marginal totals in Table 2 are fixed, the following two quantities are the mean and variance of hypergeometric distribution under the null hypothesis that the odds ratio is one.

$$E_i = \frac{(x_{Ti} + x_{Ci})n_{Ti}}{N_i}, \tag{10}$$

and

$$V_i = \frac{(x_{Ti} + x_{Ci})(N_i - x_{Ti} - x_{Ci})n_{Ti}n_{Ci}}{N_i^2(N_i - 1)}. \tag{11}$$

Based on E_i and V_i, the Peto estimate of pooled log odds ratio from K independent tables has the following expression:

$$ln(\widehat{OR})_{Peto} = \frac{\sum_{i=1}^{K}(x_{Ti} - E_i)}{\sum_{i-1}^{K} V_i}, \tag{12}$$

and

$$Var[ln(\widehat{OR})_{Peto}] = \frac{1}{\sum_{i-1}^{K} V_i}. \tag{13}$$

The Peto estimator of the combined odds ratio is not a consistent estimator and can provide severely biased results when applied to unbalanced data Greenland and Salvan (1990). The validity of the Peto estimator in a meta-analysis of rare event studies is limited to the analysis of reasonably balanced studies that have odds ratio close to 1.

3.4. Arcsine transformation

The arcsine transformation method is needed when the objective is to include all studies in the meta-analysis, including those with very rare events or zero total events, while stabilizing variance estimates to provide more accurate intervention effect estimates. Zero event in either or both arms of a given study/trial does not necessarily indicate that the true probability of an event is 0. On the contrary, it indicates that the event probability is very small, and the sample size in the study is not large enough to observe an event. The arcsine transform method estimates the combined effect by combining all studies including the zero total event studies. The arcsine difference (AS) measure of intervention effect for the ith study is defined as:

$$AS_i = arcsin\sqrt{p_{Ti}} - arcsin\sqrt{p_{Ci}}, \tag{14}$$

and its asymptotic variance given in equation (15) is finite and non-zero and depends only on the study sample size.

$$\sigma^2_{AS_i} = \frac{1}{4n_{Ti}} + \frac{1}{4n_{Ci}}. \tag{15}$$

Similar to the MH and Peto methods, the combined AS is a weighted mean of the individual AS_i's, where the $w_i = 1/\sigma^2_{AS_i}$ are the weights. Therefore,

$$\widehat{AS} = \frac{\sum_{i=1}^{K} w_i AS_i}{\sum_{i=1}^{K} w_i}. \tag{16}$$

Rücker *et al.* (2009) has recommended using $0.42/n$ instead of $1/4n$ in equation (15) to estimate the variance conservatively for small event probabilities. Simulation studies of Rücker *et al.* (2009) suggest that the bias of the estimate is slightly higher than the other two methods mentioned above. The key advantage of this method is the variance stabilizing property of the arcsine transformation, which leads to more robust estimation, even for the rare events Rücker *et al.* (2009). Nevertheless, a lack of direct interpretation has limited its wider use as a measure of intervention effect.

3.5. Heterogeneity

Fixed-effect methods like Mantel-Haenszel and Peto can significantly overestimate treatment effects in meta-analysis of rare events, especially in the presence of heterogeneity, leading to inflated type I error rates and biases. Historically, the treatment effect heterogeneity has not received sufficient attention in the context of meta-analysis of rare events. Using a continuity correction, majority of simulation studies are performed under the assumption of fixed treatment effect Sweeting *et al.* (2004); Bradburn *et al.* (2007), or a small heterogeneity Rücker *et al.* (2009). The rationale behind selecting the fixed treatment effect is the negligible heterogeneity. The zero estimate, however, is not always due to the absence of heterogeneous treatment effects but mainly due to the unavailability of adequate methods. On the other hand, studies that evaluate heterogeneity are conducted for moderate event rates. As a result, the performance of fixed-effect methods in the presence of heterogeneity is not well understood, particularly for low event rates. Although those methods are expected to perform poorly, only a few studies have extensively explored specific characteristics of the poor performance. For example, Bhaumik *et al.* (2012) showed that the asymptotic bias of combined odds ratio (with constant continuity correction "a") in the presence of treatment heterogeneity to be

$$\begin{aligned} Bias(\hat{\theta}_{wa}) = &-\frac{(p_{t|\epsilon} - q_{t|\epsilon})}{n(q_{t|\epsilon}p_{t|\epsilon})}\left\{a + \frac{p_c q_c - p_{t|\epsilon}q_{t|\epsilon}}{2(p_{t|\epsilon}q_{t|\epsilon} + p_c q_c)}\right\} \\ &+ \frac{(p_c - q_c)}{n(p_c q_c)}\left\{a + \frac{p_{t|\epsilon}q_{t|\epsilon} - p_c q_c}{2(p_{t|\epsilon}q_{t|\epsilon} + p_c q_c)}\right\}, \end{aligned} \tag{17}$$

where $p_{t|\epsilon}$ and p_c are unobservable underlying true event rates, $\hat{\theta}_{wa} = \sum_{i=1}^{k} \hat{w}_i(\tau^2)\hat{\theta}_{ia} / \sum_{i=1}^{k} \hat{w}_i(\tau^2)$, and $\hat{w}_i(\tau^2) = \dfrac{1}{\hat{\sigma}_i^2(\tau^2)}$. Their simulation study suggests that, for low event rates, the Mantel-Haenszel and Peto methods can grossly overestimate the treatment effect (see Figure 1) and produce an unacceptably high type I error rate. Therefore, in the presence of heterogeneity, the behavior of fixed-effect methods does not follow the patterns demonstrated in the simulation studies without heterogeneity. The bias of the treatment effect is reduced when random-effects methods with 1/2 continuity correction are used along with the improved estimates of heterogeneity parameters. However, even with alternative methods, an estimate of heterogeneity may not produce a non-zero value when event rates are extremely low (*e.g.*, 1/1000). As the true state of heterogeneity is unknown a priori, a large bias and an inflated type I error rate (see Figure 1) are potential threats associated with the validity of estimates of treatment effects obtained from weighted average methods, including Mantel-Henszel, Peto, and Dersimonial-Liard. These undesirable characteristics become more pronounced for low event rates. Shuster (2010) has also raised similar concerns regarding the validity of empirically based weighting in random effects and demonstrated that empirical weighting produces substantial bias for the DerSimonian-Laird approach.

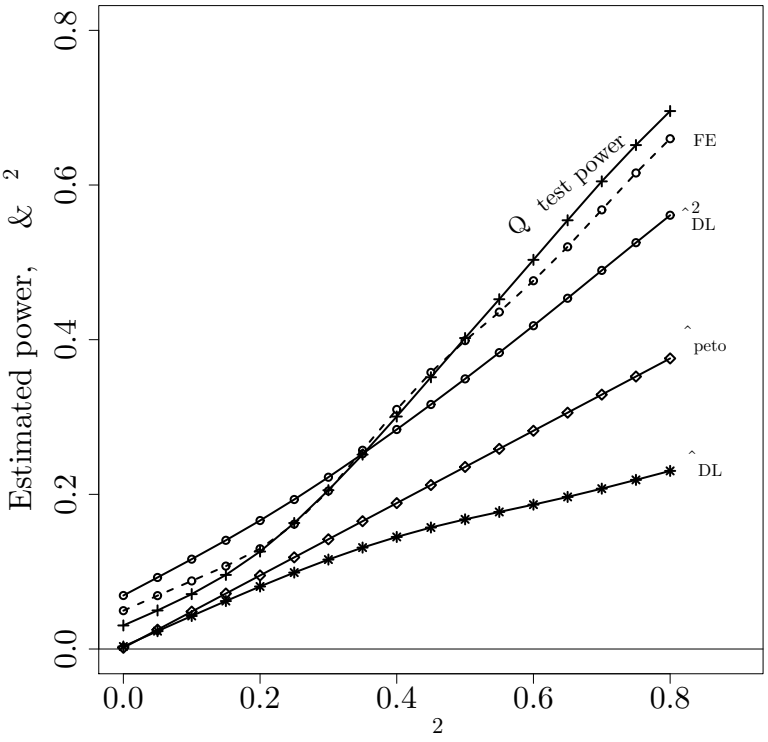


Figure 1: Power curve of Q-test as a function of τ^2 for a low event rate (0.4%). The true value of θ is set at 0. The α_{FE} is a type I error rate for testing the null hypothesis using the fixed-effect (Peto) method.

4. Likelihood based methods

4.1. Maximum marginal likelihood methods

The maximum marginal likelihood (MML) method in meta-analysis allows for simultaneous estimation of treatment effects and heterogeneity parameters, accommodating studies with zero total events without requiring continuity corrections. The MML method is model-based, an alternative to the moments matching methods. The major advantage of the MML approach over traditional methods is that the zero total events studies can be included without any artificial continuity corrections. It does have the flexibility of estimating both the overall treatment effect, and the heterogeneity parameter(s) simultaneously.

Consider an observed 2×2 Table 2 for the ith study for a meta-analysis of k studies. Suppose the probability of observing an event in the ith study is p_{ti} for the treatment group and p_{ci} for the control group. The log-odds of adverse events in group $j \in \{T, C\}$ can be modeled as follows.

$$ln\left(\frac{p_{ji}}{1-p_{ji}}\right) = \mu + \epsilon_{1i} + (\theta + \epsilon_{2i})T_{ji} \tag{18}$$

where T_{ji} is the treatment indicator variable defined as $T_{ji} = 0$ for $j = c$ and $T_{ji} = 1$ for $j = t$; and $\epsilon_1 \sim N(0, \sigma_\mu^2)$ and $\epsilon_2 \sim N(0, \tau^2)$ are the random-effects associated with mean

log-odds of an event in control group μ, and treatment effect θ, such that

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 & \rho\sigma_\mu\tau \\ \rho\sigma_\mu\tau & \tau^2 \end{pmatrix} \right\}. \tag{19}$$

Therefore, this model allows for heterogeneity in both the baseline risk and the treatment effect. Conditional on the random effects, the likelihood function for the ith study is

$$l(\boldsymbol{x_i}|\epsilon) = p_{ti}^{x_{ti}} q_{ti}^{n_{ti}-x_{ti}} p_{ci}^{x_{ci}} q_{ci}^{n_{ci}-x_{ci}}, \tag{20}$$

where $\boldsymbol{x_i} = (x_{ti}, x_{ci})$ is the vector pattern of responses from study i. The models (18)-(19) involve three parameters μ, θ, and Σ, where Σ denotes covariance matrix on the right hand side of Equation (19). The marginal likelihood function for these parameters is obtained by integrating the conditional likelihood (20) over the distribution of random effects as follows

$$h(\boldsymbol{\beta}; \boldsymbol{x_i}) = h(\boldsymbol{x_i}) = \int_\epsilon l(\boldsymbol{x_i}|\epsilon) g(\epsilon) d\epsilon, \tag{21}$$

where $g(\epsilon)$ represents the related bivariate normal density. As studies are assumed to be independent, the full log-likelihood for k studies can be expressed as

$$\log L = \sum_{i=1}^{k} \log h(\boldsymbol{x_i}), \tag{22}$$

and for a parameter vector $\boldsymbol{\beta} = (\mu, \theta, \Sigma)$, the first derivatives of the log-likelihood with respect to $\boldsymbol{\beta}$ are

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{k} \frac{1}{h(\boldsymbol{x_i})} \frac{\partial h(\boldsymbol{x_i})}{\partial \boldsymbol{\beta}}, \tag{23}$$

where

$$\frac{\partial h(\boldsymbol{x_i})}{\partial \boldsymbol{\beta}} = \int_\epsilon \frac{\partial \log l(\boldsymbol{x_i}|\epsilon)}{\partial \boldsymbol{\beta}} l(\boldsymbol{x_i}|\epsilon) g(\epsilon) d\epsilon \,. \tag{24}$$

A close-form solution of (24) is generally not available for nonlinear models. Therefore, numerical techniques such as Gauss-Hermite quadrature are required for the integration of the random effect space (i.e., ϵ). The marginal likelihood equation in (21) can be approximated numerically to any practical degree of accuracy by summing on a specified number of quadrature nodes and the corresponding quadrature weights. Commercial software packages such as SAS, STATA, SuperMix . can easily fit MML models, and the GLIMMIX procedure in SAS or the glmer package in R can be used to fit alternative linearized approximation to (24).

The MML models offer a variety of modeling strategies in the context of meta-analysis. Treatment effect may be estimated with a single random effect (background incidence or treatment effect) or a model with two correlated random effects. However, this flexibility to construct a model with a combination of multiple random effects also creates room for model mis-specifications. The detailed analysis of the impact of such model misspecification on the characteristics and testing of the overall effect estimator and the heterogeneity parameter has shown that the models that allow heterogeneity in both baseline rate and treatment effect

across studies have low type I and type II error rates, and are the least biased compared to other model specifications Amatya *et al.* (2015).

4.2. Beta-binomial model

The beta-binomial model offers a Bayesian framework for meta-analysis, allowing for estimation of treatment effects and correlations between event probabilities across studies. The beta-binomial model is another alternative to the moment-based methods. In the Bayesian setup, a meta-analysis of binary events can be performed in two ways using the beta-binomial model. The first way is to adopt a univariate approach, where event probabilities p_{Ti} and p_{Ci} are assumed to be independent. However, the individual binary observations within the jth arm of the ith study (which add up to x_{ji} for $j =\in \{T, C\}$) are allowed to be correlated by imposing $p_{ji} \sim beta(\alpha_j, \beta_j)$ as a prior. As a result, $E(p_j) = \mu_j = \frac{\alpha_j}{\alpha_j+\beta_j}$, $Var(p_j) = \mu(1-\mu)\theta/(1+\theta)$ with $\theta = 1/(\alpha_j + \beta_j)$, and the correlation between observations within jth arm of each study is $\rho_j = 1/(\alpha_j + \beta_j + 1)$, and the marginal distribution of x_{ji} is the beta-binomial distribution with the following log-likelihood function:

$$\begin{aligned} l_{ji}(\alpha_j, \beta_j) =& ln\Gamma(n_{ji}+1) + ln\Gamma(x_{ji}+\alpha_j) + ln\Gamma(n_{ji}-x_{ji}+\beta_j) \\ &+ ln\Gamma(\alpha_j+\beta_j) - ln\Gamma(x_{ji}+1) - ln\Gamma(n_{ji}-y_{ji}+1) \\ &- ln\Gamma(n_{ji}+\alpha_j+\beta_j) - ln\Gamma(\alpha_j) - ln\Gamma(\beta_j), \end{aligned} \tag{25}$$

and the joint log-likelihood function is:

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{K} \sum_{j\in\{T,C\}} l_{ji}(\alpha_j, \beta_j). \tag{26}$$

The number of parameters is reduced further by modeling the mean function $g(\mu_j) = b_0 + b_1 x_j$, where g is a link function as in the generalized linear model, and $x_j = 1$, if $j = T$ and $x_j = 0$, if $j = C$. A specific link function for g determines the type of effect. For example, the logit link function gives the log odds ratio, and the log link function measures log relative risk. Kuss (2014) recommends avoiding the identity link to estimate the risk difference and suggests to use the estimated event probabilities $\hat{p}_C = g^{-1}(\hat{b}_0)$ and $\hat{p}_T = g^{-1}(\hat{b}_0 + \hat{b}_1)$ from the logit model for the control and treatment groups, respectively.

The second approach is to use the bivariate beta-binomial model which addresses the correlation between the event probabilities of two treatment arms of the studies. The correlation between control event rates (proportion) and treatment effects has been identified in studies by various authors (Schmid *et al.*, 1998, and references therein). Unlike the MML, the bivariate beta-binomial model implies a linear relationship between p_T and p_C on the original scale. Chu *et al.* (2012) described a beta-binomial model in two stages. In the first stage, X_{ji} is assumed to be independently binomially distributed, such that

$$P(X_{Ti} = x_{Ti}, X_{Ci} = x_{Ci} | n_{Ti}, n_{Ci}, p_{Ti}, p_{Ci}) = \prod_{j\in\{T,C\}} \binom{n_{ji}}{x_{ji}} (p_{ji})^{x_{ji}} (1-p_{ji})^{n_{ji}-x_{ji}}. \tag{27}$$

In the second stage, the joint distribution of p_{Ti}, and p_{Ci} is specified using a Sarmanov beta prior distribution as follows (see Luo *et al.*, 2014):

$$\begin{aligned} p_{Ti}, p_{Ci}|\alpha_T, \alpha_C, \beta_T, \beta_C &\sim f(p_T, p_C; \alpha_T, \alpha_C, \beta_T, \beta_C) \\ &= beta(p_T; \alpha_T, \beta_T) beta(p_C; \alpha_C, \beta_C) \left(1 + \rho \frac{(p_T - \mu_T)(P_C - \mu_C)}{\delta_T \delta_C}\right), \end{aligned} \tag{28}$$

where ρ is the correlation coefficient between p_{Ti} and p_{Ci}; $beta(p; \alpha, \beta) = [\mathrm{B}(\alpha, \beta)]^{-1} p^{\alpha-1}(1-p)^{\beta-1}$ with $\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$; and $\mu_j = \alpha_j/(\alpha_j+\beta_j)$, $\delta_j^2 = \mu_j(1-\mu_j)/(\alpha_j+\beta_j+1)$, and $j \in \{T, C\}$. As a result, the log marginalized likelihood function for the unknown hyperparameters $(\alpha_T, \alpha_C, \beta_T, \beta_C, \rho)$ is

$$\begin{aligned} &log\ L(\alpha_T, \alpha_C, \beta_T, \beta_C, \rho) \\ &= \sum_{i=1}^{k} log\left[P_{BB}(x_{Ti}; n_{Ti}, \alpha_T, \beta_T) P_{BB}(x_{Ci}; n_{Ci}, \alpha_C, \beta_C)\right] \\ &+ log\left\{1 + \rho \frac{\left(\frac{x_{Ti}+\alpha_T}{n_{Ti}+\alpha_T+\beta_T} - \mu_T\right)\left(\frac{x_{Ci}+\alpha_C}{n_{Ci}+\alpha_C+\beta_C} - \mu_T\right)}{\delta_T \delta_C}\right\}, \end{aligned} \tag{29}$$

where $P_{BB}(x; n; \alpha; \beta)$ is the probability mass function of a beta-binomial distribution, such that

$$P_{BB}(x; n; \alpha; \beta) = \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)}. \tag{30}$$

The maximum likelihood estimates $(\hat{\alpha}_T, \hat{\alpha}_C, \hat{\beta}_T, \hat{\beta}_C, \hat{\rho})$ is obtained by maximizing likelihood function (29). Based on these estimates, three overall effect measures are estimated as follows:

$$\text{Odds Ratio} = \widehat{OR} = \frac{\hat{\mu}_T/(1-\hat{\mu}_T)}{\hat{\mu}_C/(1-\hat{\mu}_C)} = \frac{\hat{\alpha}_T \hat{\beta}_C}{\hat{\alpha}_C \hat{\beta}_T}, \tag{31}$$

$$\text{Relative Risk} = \widehat{RR} = \frac{\hat{\mu}_T}{\hat{\mu}_C} = \frac{\hat{\alpha}_T/(\hat{\alpha}_T+\hat{\beta}_T)}{\hat{\alpha}_C/(\hat{\alpha}_C+\hat{\beta}_C)}, \tag{32}$$

$$\text{Risk difference} = \widehat{RD} = \hat{\mu}_T - \hat{\mu}_C = \frac{\hat{\alpha}_T}{(\hat{\alpha}_T+\hat{\beta}_T)} - \frac{\hat{\alpha}_C}{(\hat{\alpha}_C+\hat{\beta}_C)}. \tag{33}$$

The variances of these estimates are calculated using the delta method.

5. Confidence distribution methods

Confidence distribution, in meta-analysis refers to a statistical method where the uncertainty about a parameter (such as an effect size) is represented by a distribution rather than a single point estimate. This distribution integrates information from multiple studies, accommodating varying study sizes and results, including studies with zero total events. Xie *et al.* (2011) have developed a unified framework for meta-analysis by combining confidence distributions (CD) from individual studies. The combined CD function is obtained

by appropriately weighting the individual distribution estimators. This is in contrast to the traditional meta-analysis, where a combined estimate is obtained by averaging individual point estimates with appropriate weights. The combined CD does have various optimality conditions. This method also allows straightforward integration of data from all studies including zero total events.

Suppose that the CD function $H_i(\theta) = H_i(\boldsymbol{X}_i, \theta)$, $i = 1, \ldots, k$ for the parameter θ can be obtained from each study with corresponding samples $\boldsymbol{X}_i$ of size n_i. A combined confidence distribution function across k studies (H_c) is constructed as

$$H_c = G_c\{g_c(H_1(\theta), \ldots, H_k(\theta))\}, \tag{34}$$

where $g_c(u_1, \ldots, u_k) = w_1 F_0^{-1}(u_1) + \ldots + w_k F_0^{-1}(u_k)$ is a monotonic function that has the cumulative distribution function $G_c(t) = P(g_c(U_1, \ldots, U_k) \leq t)$ for $U_i \sim U[0, 1]$. The transformation function $F_0(\cdot)$ is weighted by fixed positive weights $w_i \geq 0$. The conventional fixed- and random-effect meta-analysis approaches can be easily derived using the recipe in (34) (see Xie *et al.*, 2011).

5.1. Odds ratio

Meta-analysis of rare event studies using odds ratio under the CD framework was developed by Liu (2012). This method uses exact p-values based on mid-p adaptation of Fisher's exact test for the odds ratio as the CD functions for individual studies and combines them by applying the general CD combination method as described in (34). Using this exact test, the p-value function for the odds ratio Ψ is obtained as follows:

$$p_i(\Psi) \equiv p_i(\Psi; x_{Ti}, x_{Ti}) = Pr_\Psi(X_{Ti} > x_{Ti} | T_i = t_i) + \frac{1}{2} Pr_\Psi(X_{Ti} = x_{Ti} | T_i = t_i), \tag{35}$$

where, the hypothesis of interest is

$$H_0 : \Psi = \Psi^0 vs. H_1 > \Psi^0.$$

The X_{Ti}'s are assumed to follow a hypergeometric distribution conditional on $T_i = X_{Ti} + X_{Ci}$. Then, for $L_i = max(0, t_i - n_{Ci})$, and $U_i = min(n_{Ti}, t_i)$. It follows that

$$Pr_\Psi(X_{Ti} = x_{Ti} | T_i = t_i) = \frac{\binom{n_{Ti}}{x_{Ti}} \binom{n_{Ci}}{t_i - x_{Ti}} \Psi^{x_{Ti}}}{\sum_{s=L_i}^{U_i} \binom{n_{Ti}}{s} \binom{n_{Ci}}{t_i - s} \Psi^s}, \quad L_i \leq x_{Ti} \leq U_i. \tag{36}$$

The statistic $p_i(\Psi^0)$ asymptotically follows $U(0, 1)$. However, for the meta-analysis of rare events, the asymptotic conditions are seldom valid, causing a substantial deviation of $p_i(\Psi^0)$ from $U(0, 1)$. Nonetheless, Liu (2012) has shown that the general idea of a CD combining algorithm can still be used in the finite sample setting after some adjustments. They also showed that zero total event studies can provide meaningful contributions in the presence of uncertainty. The impact of zero total event studies is appropriately accounted for in the

sample size computation of the corresponding studies by using the weights:

$$w_i \propto \left[\{n_{Ti}\pi_{Ti}(1-\pi_{Ti})\}^{-1} + \{n_{Ci}\pi_{Ci}(1-\pi_{Ci})\}^{-1}\right]^{-1/2}, \tag{37}$$

which requires estimates of π_{Ci} and π_{Ti}. To improve an efficiency of the overall estimate, Liu *et al.* (2014) proposed to model π_{Ci} using a beta(β_1, β_2) distribution. The parameters of this beta distribution are estimated as follows:

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\Psi}) = \underset{\beta_1,\beta_2,\Psi}{\arg\max} \sum_{i=1}^{k} Log \int_0^1 f_\psi(x_{Ci}, x_{Ti}|\pi_{Ci}) f_{\beta_1,\beta_2}(\pi_{Ci}) d\pi_{Ci}, \tag{38}$$

where $f_{\beta_1,\beta_2}(\pi_{Ci}) = \pi_{Ci}^{\beta_1-1}(1-\pi_{Ci})^{\beta_2-1} / \int_1^0 \pi_{Ci}^{\beta_1-1}(1-\pi_{Ci})^{\beta_2-1} d(x_{Ci})$, $f_\psi(x_{Ci}, x_{Ti}|\pi_{Ci}) = c(x_{Ci}, x_{Ti})\pi_{Ci}^{x_{Ci}}(1-\pi_{Ci})^{n_{Ci}-x_{Ci}}\pi_{Ti}^{x_{Ti}}(1-\pi_{Ti})^{n_{Ti}-x_{Ti}}$, and $\pi_{Ti} = (\Psi\pi_{Ci})/(1-\pi_{Ci}+\Psi\pi_{Ci})$. The mean of the empirical conditional density of π_{Ci} is used as an estimate of π_{Ci} and an estimate of π_{Ti} is calculated through $\hat{\pi}_{Ti} = (\Psi\pi_{Ci})/(1-\hat{\pi}_{Ci}+\hat{\Psi}\hat{\pi}_{Ti})$. This manipulation produces positive estimates of π_{Ti} and π_{Ci} even for zero total event studies, allowing the inclusion of these studies without any continuity correction. When $x_{Ti} = 0$ for all i, limiting weights are calculated as follows

$$\lim_{\hat{\Psi}\to 0} \left(w_i / \sum_{i=1}^{k} w_i \right)^2 = \frac{n_{Ci}x_{Ci}/(1-x_{Ci})}{\sum_{i=1}^{k} n_{Ci}x_{Ci}/(1-x_{Ci})}.$$

The case where $x_{Ci} = 0$ for all i is handled similarly.

5.2. Risk difference

Tian *et al.* (2009) proposed a simple procedure to construct a $100(1-\alpha)$ 1-sided confidence interval (CI) of the type (a, ∞) for a common risk difference parameter Δ, based on all data from k independent studies without any artificial continuity correction. Suppose that n sets of k study-specific 1-sided CIs of any arbitrary level η can be constructed for Δ. Let $J_{ij} = (a_{ij}, \infty)$ be the η_j-level 1-sided CI obtained from the ith study, for $i = 1, \ldots, k$, and $j = 1, \ldots, n$; such that $0 < \eta_1 < \eta_2 < \ldots < \eta_n < 1$, and $a_{i1} > a_{i2} > \ldots > a_{in}$. The final combined interval for δ is (see Tian *et al.*, 2009)

$$\sum_{i=1}^{k} w_i \sum_{j=1}^{n} \tilde{w}_j \{(I(\Delta > a_{ij}) - \eta_j\} \geq c, \tag{39}$$

where $I(\cdot)$ is the indicator function, w_i is a study-specific weight, $\tilde{w}_j$ is a positive weight for η_j-level intervals, and the critical value c is chosen such that

$$Pr\left[\sum_{i=1}^{k} w_i \sum_{j=1}^{n} \tilde{w}_j (B_{ij} - \eta_j) < c\right] \leq \alpha. \tag{40}$$

In equation (40), $(B_{i1}, \ldots, B_{ik})$ are n independent random vectors whose components are correlated Bernoulli variables such that $B_{i1} \leq B_{i2} \leq \ldots \leq B_{ik}$ and $pr(B_{ij} = 1) = \eta_j$.

Tian *et al.* (2009) suggested to use $w_i = 1/(n_{Ti}+n_{Ci})$, and $\tilde{w}_j = \{\eta_j(1-\eta_j)\}^{-1}$ for the weights. Yang *et al.* (2012) showed this procedure to be a special case under the CD framework, where $F_0^{-1}(u)$ is chosen to be $\sum_{j=1}^{n} \tilde{w}_j \{I(u > 1-\eta_j) - \eta_j\}$. Then,

$$H^c(\Delta) = G_c \left\{ \sum_{i=1}^{k} w_i \sum_{j=1}^{n} \tilde{w}_j \{I(H_i(\delta) > 1-\eta_j) - \eta_j\} \right\}. \tag{41}$$

For the detailed derivation of the proof, see Yang *et al.* (2012), where alternative equivalent expression for $H^c(\Delta)$ is also obtained by using the logistic function as a transformation function.

6. Illustration

In their highly influential meta-analysis article, Nissen and Wolski (2007) concluded that rosiglitazone was associated with a significant risk of myocardial infarction [odds ratio (OR) 1.43, 95 % CI (1.03,1.98), P = 0.03] and an increase in the risk of death from cardiovascular causes, which had borderline significance [OR 1.64, 95 % CI (0.98 2.74); P = 0.06]. These conclusions were based on a fixed-effect meta-analysis using the Peto method. Soon after the release of these results, a series of reanalysis of the same data was published by others using different methods. Diamond *et al.* (2007) has conducted the meta-analysis using three conventional fixed-effect methods with two continuity corrections and including/excluding zero total event studies. Stoto (2015) reported some results based on a Localio *et al.* (2008) wide variety of statistical methods. Tian *et al.* (2009), Chu *et al.* (2012), and Liu *et al.* (2014) have used a few relatively new approaches to analyze the rosiglitazone data. They included all studies (including zero total event studies) without any continuity correction. Chu *et al.* (2012) used the beta-binomial model, whereas Tian *et al.* (2009), and Liu *et al.* (2014) used the confidence distribution methods. Estimates of various effect measures from these articles are summarized in Table 3.

7. Bayesian methods

The Bayesian methodology in meta-analysis offers flexible modeling with hierarchical structures, integrating prior information and accommodating non-normal distributions of random effects. Computational intensity has decreased with advancements in Monte Carlo techniques and computing power, supporting complex analyses without major barriers. Bayesian methodology is an alternative to the traditional meta-analysis methods. It provides a broad range of modeling alternatives with multiple levels of hierarchy and naturally integrates prior information on parameters of interest from other trials or studies. The emphasis on hierarchical modeling accounts for uncertainty in all parameters including the between-study heterogeneity. The flexibility of the Bayesian approach allows for rigorous sensitivity analysis, which is particularly important for meta-analysis of rare events. Furthermore, the Bayesian framework can be easily extended to non-normal distributions of random effects. The computational complexity of the Bayesian approach is substantially intensive compared to the traditional methods. Fortunately, software is readily available that incorporates rapidly developing Monte Carlo techniques. Due to the unprecedented rise in computational power of modern personal computers, complex computation in Bayesian analysis is no longer a major barrier.

Table 3: Various estimates of effect measures for rosiglitazone MI data

	Method	CC	OR (95% CI)	RR (SE)	RD (95% CI)
Nissen	Peto	0.5	1.43 (1.03, 1.98)		
Diamond	Fixed, IV	TAC	1.34 (.097, 1.84)		
	Fixed, IV	CC	1.29 (0.94, 1.76)		.0015 (0, .0031)
	Fixed, MH	TAC	1.36 (1.00, 1.84)		
	Fixed, MH	CC	1.28 (0.95, 1.72)		.0020 (0, .0041)
	Fixed, MH	TAC+	1.35 (1.00, 1.82)		
	Fixed, MH	CC+	1.26 (0.93, 1.69)		
Localio	Random [DL]	NA	1.31 (0.91, 1.89)		
	Random [DL]	0.5	1.31 (0.95, 1.79)		
	Random [DL]	TAC	1.33 (0.93, 1.91)		
	Conditional logistic	NA	1.45 (1.05, 2.01)		
	Exact stratified	NA	1.45 (1.03, 2.04)		
	Random intercept/slope	NA	1.37 (0.99, 1.90)		
Chu	Bivariate beta-binomial	NA		1.291 (0.382)	0.0011 (SE=0.0013)
Tian	Exact CD	NA			0.0018 (-0.008, 0.004)
Liu	Exact CD	NA	(.972, 2.00)		
	Adjusted Exact CD		(1.04, 2.01)		

CC: Constant (0.5) correction for continuity, CC+: constant correction for continuity that includes all zero total event studies, IV: inverse variance, MH: Mantel-Haenszel, TAC: treatment arm correction for continuity, TAC: treatment arm correction for continuity that includes all zero-total-event studies

The key elements of a generic Bayesian meta-analysis model are the prior distributions on both the effect and the heterogeneity parameters. The simplest form of Bayesian random-effect meta-analysis is as follows: (see Sutton and Abrams, 2001):

$$\begin{aligned} \hat{\theta}_i &\sim f(\hat{\theta}_i|\theta_i, \sigma_i^2) \\ \theta_i &\sim \pi(\theta_i|\theta, \tau^2) \\ \theta &\sim h(\theta) \\ \tau^2 &\sim h(\tau^2), \end{aligned} \tag{42}$$

where $h(\theta)$ and $h(\tau^2)$ are the prior distributions of effect parameter θ, and between-study heterogeneity parameter τ^2 respectively. The resulting posterior distribution does have the following form:

$$p(\theta, \tau, \theta_i|\hat{\boldsymbol{\theta}}_i) \propto h(\theta)h(\tau^2) \prod_{i=1}^{K} \pi(\theta_i|\theta, \tau^2) \prod_{i=1}^{K} f(\hat{\theta}_i|\theta_i, \sigma_i^2). \tag{43}$$

Inferences on parameters of interest are made from the mode of the posterior distribution (43). Except for some cases of conjugate prior distributions, the posterior mode is usually

not available in its closed form. Instead, Monte Carlo methods such as Gibbs sampling are used to numerically approximate the mode of the posterior distribution. An example of Gibbs sampling is found in a meta-analysis of randomized controlled trials comparing sodium monobuorophosphate (SMFP) to sodium Buoride (NaF) dentifrices (toothpaste) in the prevention of caries development Abrams and Sanso (1998). A complex example of a Bayesian hierarchical model that incorporates a study-level component of variability and facilitates extensive sensitivity analysis is found in Kaizar *et al.* (2006).

7.1. Strong prior

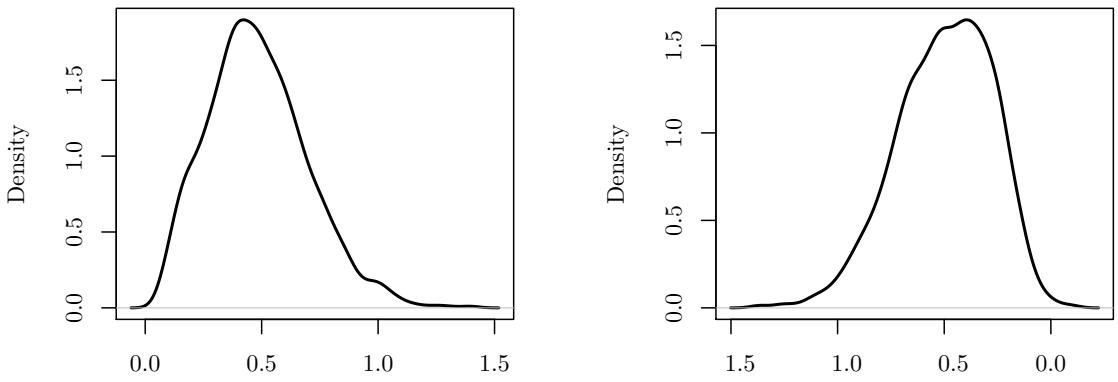
A strong prior in Bayesian analysis is one that conveys substantial prior belief or information about the parameters of interest, such as the probability of events in the control arm, treatment effect, or between-study heterogeneity. It can significantly influence the estimated outcomes of a meta-analysis by anchoring the inference towards specific values based on empirical data or subjective judgment. The prior distribution is not only a key part of Bayesian analysis, but also it is one of the most difficult and controversial aspects of the analysis. A non-informative prior is specified to express vague or general information of a parameter and to minimize a perceived subjective bias. On the other hand, an alternative prior distribution can be specified to integrate prior belief or substantiated information relevant to the estimation of the parameter of interest. Such informative priors may be formulated by considering the plausible range of the parameters, based on observed distributions from empirical studies, or based purely on subjective clinical judgment Warn *et al.* (2002). These informative priors may influence the conclusion of the meta-analysis. When binary events are of concern, prior distribution needs to be specified for the following three parameters:(1) probability of events in the control arm, (2) treatment effect, and (3) between-study heterogeneity. Strong priors on some of these parameters may have a substantial impact on the estimated overall treatment effect. The following Bayesian meta-analysis of rositaglitazone data illustrates the impact of strong priors.

Let n_{ji} be the number of the participants in the ith trial who received the jth treatment. Suppose that the probability of experiencing MI is p_{ji}. The observed MI incidences x_{ji} may be modeled under the Bayesian framework as follows:

$$\begin{aligned} X_{ji} &\sim binomial(p_{ji}, n_{ji}), \quad \text{for } j \in \{T, C\}, \text{ and } i = 1, 2, \ldots, k \\ \theta_i &\sim N(\theta, \tau^2), \quad \mu_i = logit(p_{Ci}) \\ logit(p_{Ti}) &= \mu_i + \theta_i. \end{aligned}$$

The model presented above can be easily implemented in the WinBugs program (see Warn *et al.*, 2002). To represent a plausible range of θ and τ, prior distributions $N(0, 10)$, and $U(0, 2)$ are specified, respectively. Based on the specified priors and the observed data, the WinBugs program computes posterior distributions of parameters using MCMC methods. The posterior modes of θ and τ are estimated from these posterior distributions. A graphical representation of the posterior distribution for this example data is displayed in Figure (2).

The posterior estimate of the combined odds ratio and heterogeneity from the above modeling are given in Table 4, where noninformative independent prior $U(0, 1)$ is specified for P_{Ci}. This model specification assumes a fixed background MI incidence rate and heterogeneous between-study treatment effects. The results in Table 4 show a posterior estimate



(a) Posterior Distribution of τ. **(b) Posterior Distribution of θ.**

Figure 2: Posterior Distributions of τ and θ.

Table 4: Bayesian Meta-Analysis of 42 Rosiglitazone Trials

	mean	2.5%	25%	50%	75%	97.5%
OR	0.608	0.379	0.522	0.617	0.720	0.883
τ^2	0.234	0.015	0.111	0.217	0.379	0.932

of the odds ratio of 0.608 with 95% credible interval of (0.379, 0.883), which is markedly different from the moment based estimators in Table 3. It is noteworthy that the current estimate is close to the estimate obtained from MML with random effects restricted only to the treatment effect. A priori belief regarding the incidence of MI rate among type II diabetes patients can be integrated by changing the parameters of the prior distribution of P_{Ci}. Figure 3 displays the impact of different values of parameters of the prior distribution of P_{Ci}. The posterior odds ratio remains below 1.0 as the prior becomes closer to the vague, the same as the results in Table 4. However, a strong prior of uniform(0, 0.01) provides a positive log odds ratio that is close to the moment-based results. Figure 3 essentially shows that if one is willing (or has reason) to believe *a priori* that the prevalence of MI is extremely rare, *e.g.*, less than 6/1000, in a diabetic population, then the observed data supports an elevated risk of MI among rosiglitazone users. In the absence of such prior information, the model does not support the conclusion derived from the moment-based analyses.

This example clearly demonstrates the effect of prior distributions on the conclusion of meta-analysis. A similar but less dramatic effect on the estimate of the log odds ratio is also observed for different informative prior specifications of τ. However, when only a small number of studies are available, a strong prior distribution on τ can significantly influence the results of the analysis Sutton and Abrams (2001). The hierarchical Bayesian approach is used to introduce a reasonable amount of uncertainty in the prior belief regarding distributions of the model parameters. In the rosiglitazone example, a $beta(a, b)$ prior distribution may be used to model p_c, and a $gamma(s, r)$ hyper-prior may be placed on the parameters of the

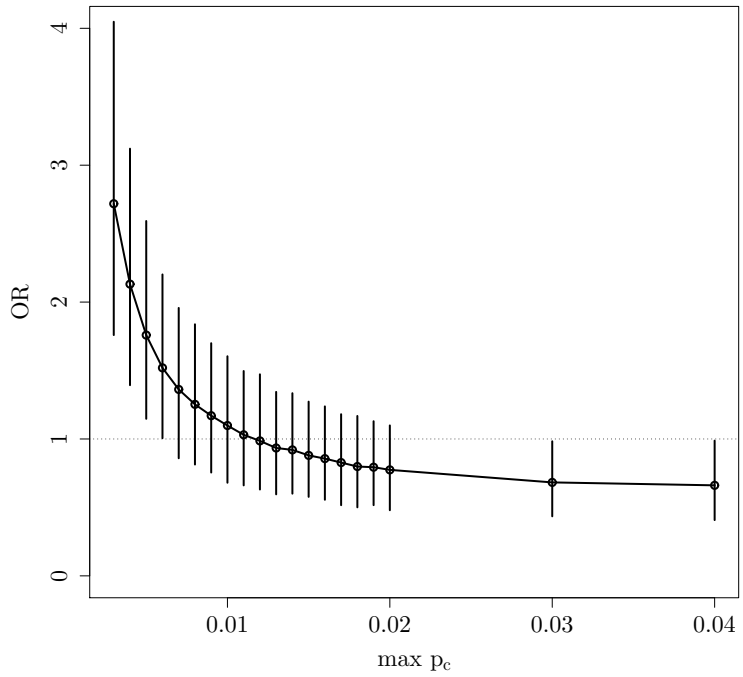


Figure 3: Posterior mean and 95% credible interval of odds ratio for varying maximum values (max p_c) of the prior distribution of $p_c \sim unif(0, \max p_c)$

beta distribution. The resulting full Bayesian model is

$$\begin{aligned} &r_{Ti} \sim binomial(p_{Ti}, n_{Ti}),\ r_{Ci} \sim binomial(p_{Ci}, n_{Ci}) \\ &\mu_i = logit(p_{Ci}),\ logit(p_{Ti}) = \mu_i + \theta_i, \\ &p_{Ci} \sim beta(a, b),\ \theta_i \sim N(\theta, \tau^2) \\ &a \sim gamma(r_a, s_a),\ b \sim gamma(r_b, s_b); \end{aligned} \tag{44}$$

where, gamma $(r, s) = \frac{s^r}{\Gamma(r)} x^{r-1} e^{-sx}$ for $x > 0, r > 0$ and $s > 0$. Table 5 presents the estimates obtained from this model for different combinations of parameters of the gamma hyper-prior distribution. For this illustration, r_a and r_b were varied while holding s_a and s_b values fixed at 0.25 and 1.5, respectively. The posterior means of the log odds ratio are clearly more consistent between 0.11 (OR=1.12) and 0.17 (OR=1.19) over different specifications of hyper-prior distributions. These estimates of log odds ratios are closer to the estimates obtained from the moment-based methods.

Table 5: Hierarchical Bayesian Analysis of 42 Rosiglitazone Trials using Different Gamma Hyper-prior Distributions.

		α	β	log odds ratio (θ)			τ^2		
r_a	r_b	mean	mean	mean	2.5%	97.5%	mean	2.5%	97.5%
0.0004	12.052	0.305	14.537	0.152	-0.252	0.513	0.069	0.000	0.651
0.0018	15.801	0.322	16.970	0.135	-0.263	0.530	0.077	0.000	0.709
0.0031	4.170	0.263	8.861	0.144	-0.206	0.529	0.062	0.000	0.600
0.0039	15.551	0.320	16.742	0.152	-0.240	0.524	0.071	0.000	0.612
0.0042	19.110	0.336	19.199	0.153	-0.251	0.551	0.065	0.000	0.622
0.0047	17.088	0.323	17.972	0.159	-0.237	0.512	0.066	0.000	0.621
0.0057	17.625	0.327	18.191	0.143	-0.223	0.549	0.061	0.000	0.627
0.0066	8.534	0.285	11.883	0.163	-0.221	0.538	0.074	0.000	0.660
0.0088	11.785	0.298	12.304	0.150	0.025	0.237	0.002	0.000	0.008
0.5884	16.796	0.332	17.959	0.146	-0.214	0.531	0.069	0.000	0.602
1.1903	7.241	0.291	11.447	0.139	-0.267	0.509	0.075	0.000	0.670
1.2831	13.824	0.317	15.963	0.112	-0.249	0.465	0.107	0.004	0.719
1.9515	1.021	0.259	7.270	0.173	-0.218	0.552	0.060	0.000	0.626
2.5283	14.976	0.342	16.910	0.137	-0.251	0.524	0.075	0.000	0.638
2.7945	16.991	0.360	18.552	0.120	-0.298	0.489	0.075	0.000	0.598
3.5064	3.797	0.292	9.580	0.161	-0.248	0.587	0.060	0.000	0.584
3.8382	15.027	0.357	17.441	0.131	-0.252	0.503	0.063	0.000	0.516
4.0664	8.528	0.326	12.920	0.123	-0.271	0.515	0.071	0.000	0.653
4.0838	7.780	0.323	12.488	0.140	-0.266	0.494	0.067	0.000	0.652

8. Discussion

Meta-analysis of safety data, particularly for rare events, poses challenges due to low event rates in randomized controlled trials (RCTs) designed primarily for efficacy. These issues include inadequate power to detect true risks and complexities arising from biases and study design differences in observational studies. Analytical methods vary in handling heterogeneity, influencing conclusions on drug safety, as seen in meta-analyses of Rosiglitazone's association with myocardial infarction, highlighting the need for cautious interpretation and sensitivity analysis. Meta-analysis of rare events data in general, and safety data in particular is a complex statistical problem with immense practical importance. Randomized control trials (RCT) are generally not designed to study safety issues related to a treatment. Therefore, individual trials may not provide adequate power to detect the true risk of adverse events, particularly when the adverse event is rare. Post-marketing safety studies are usually conducted using large observational studies. A meta-analysis from a series of large observational studies can provide a spurious degree of statistical precision, leading to acceptance of low-level associations resulting from residual confounding Henry and Hill (1999). Inherent biases and differences in study designs add further complexities to the meta-analysis of observational studies. Consequently, the assessment of drug safety partly relies on the meta-analysis of RCTs and other published literature. Although such reliance on meta-analysis holds promises of synthesizing all available evidence, it is not without serious pitfalls. Stoto (2015) discussed these issues using three high-profile examples. Stoto (2015) concluded that the precision of the results of one meta-analysis can be deceptively

low due to some typical characteristics of safety data extracted from efficacy studies. Those characteristics include low adverse event rates, untestable clinical and methodological heterogeneity, and incomplete and inconsistent reporting of adverse effects. Consequently, different syntheses can provide qualitatively different conclusions. For example, analytical methods that avoid or deal with heterogeneity in different ways may lead to different conclusions related to the risk of adverse events. A careful consideration is particularly important for safety studies, where the standard Cochrans' Q-test for detecting heterogeneity is known to be significantly underpowered (see Figure 1). These studies often possess substantive heterogeneity of the populations under study, comparison groups, and length of follow-up. The rationale for using the Peto method in such situations often points to its greater statistical power which is considered to be more important in safety analysis than the consideration of heterogeneity. However, one must not overlook a high type I error rate associated with such methods in the presence of heterogeneity. Discrepancies originating from the use of various methods are evident in the comparison of meta-analytical investigations of MI associated with rosiglitazone. A decision to place severe restrictions on the utilization of the drug was highly influenced by the results of the Peto method-based meta-analysis performed by Nissen and Wolski (2007). That analysis yielded a 95% confidence interval of (1.031, 1.979) and a p-value of 0.032 for testing that the odds ratio is 1, and thus concluded that rosiglitazone was significantly associated with myocardial infarction. The subsequent meta-analyses by others using different methods produced results that did not agree with Nissen and Wolski (2007). The varying conclusions depended on the inclusion or exclusion of zero total event studies Liu *et al.* (2014), continuity correction strategies Diamond *et al.* (2007), and effect measure (RR vs. OR) and statistical method used for analysis Stoto (2015). Furthermore, meta-analysis is itself an observational study of studies. When only a small number of adverse events are observed, meta-analysis may not be able to disentangle confounding by the indication and drug type. Over-reliance on a single analysis is not recommended when analyzing safety data. Fortunately, there are several commercial (SAS, STATA, StatXact) and freely available software (RevMan, and Rgmeta, meta, exactmeta) to facilitate an extensive sensitivity analysis when analyzing safety data involving adverse events that might occur in one per thousand patients or fewer.

References

Abrams, K. and Sanso, B. (1998). Approximate Bayesian inference for random effects meta-analysis. *Statistics in Medicine*, **17**, 201–218.

Amatya, A., Bhaumik, D. K., Normand, S.-L., Greenhouse, J., Kaizar, E., Neelon, B., and Gibbons, R. D. (2015). Likelihood-based random effect meta-analysis of binary events. *Journal of Biopharmaceutical Statistics*, **25**, 984–1004.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society*, **32**, 283–301.

Bhaumik, D. K., Amatya, A., Normand, S. T., Greenhouse, J., Kaizar, E., Neelon, B., and Gibbons, R. D. (2012). Meta-analysis of rare binary adverse event data. *Journal of the American Statistical Association*, **107**, 555–567.

Bradburn, M. J., Deeks, J. J., Berlin, J. A., and Russell L., A. (2007). Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine*, **26**, 53–77.

Chu, H., Nie, L., Chen, Y., Huang, Y., and Sun, W. (2012). Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative risk. *Statistical Methods in Medical Research*, **21**, 621–633. Epub 2010 Dec 21.

Cox, D. (1970). The continuity correction. *Biometrika*, **57**, 217–219.

Diamond, G. A., Bax, L., and Kaul, S. (2007). Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Annals of Internal Medicine*, **147**, 578–581.

Friedrich, J. O., Adhikari, N., and Beyene, J. (2007). Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC Medical Research Methodology*, **7**.

Greenland, S. and Salvan, A. (1990). Bias in the one-step method for pooling study results. *Statistics in Medicine*, **9**, 247–252.

Henry, D. and Hill, S. (1999). Meta-analysis: its role in assessing drug safety. *Pharmacoepidemiology and Drug Safety*, **8**, 167–168.

Kaizar, E., Greenhouse, J., Seltman, H., and Kelleher, K. (2006). Do antidepressants cause suicidality in children? a Bayesian meta-analysis. *Clinical Trials*, **3**, 73–98.

Kuss, O. (2014). Statistical methods for meta-analyses including information from studies without any eventsadd nothing to nothing-and succeed nevertheless. *Statistics in Medicine*, **34**, 1097–1116.

Liu, D. (2012). *Combining Information for Heterogeneous Studies and Rare Event Studies: A Confidence Distribution Approach*. PhD thesis, Rutgers University.

Liu, D., Liu, R. Y., and Xie, M. (2014). Exact meta-analysis approach for discrete data and its application to 2× 2 tables with rare events. *Journal of the American Statistical Association*, **109**, 1450–1465.

Localio, R., Cornell, J., and Mulrow, C. (2008). Much ado about avandia: the meta-analysis of rare events in the service of health policy. In: 7th International Conference on Health Policy Statistics.

Luo, S., Chen, Y., Su, X., and Chu, H. (2014). mmeta: An R package for multivariate meta-analysis. *Journal of Statistical Software*, **56**, 11.

Nissen, S. E. and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, **356**, 2457–2471.

Rücker, G., Schwarzer, G., Carpenter, J., and Olkin, I. (2009). Why add anything to nothing? the arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in Medicine*, **28**, 721–738.

Schmid, C. H., Lau, J., McIntosh, M. W., and Cappelleri, J. C. (1998). An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine*, **17**, 1923–1942.

Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine*, **29**, 1259–1265.

Stoto, M. A. (2015). Drug safety meta-analysis: Promises and pitfalls. *Drug Safety*, **38**, 233–243.

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, **10**, 277–303.

Sweeting, J. M., Sutton, A. J., and Lambert, P. C. (2004). What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, **23**, 1351–1375.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P., and Wei, L. J. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2× 2 tables with all available data but without artificial continuity correction. *Biostatistics*, **10**, 275–281.

Warn, D. E., Thompson, S. G., and Spiegelhalter, D. J. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine*, **21**, 1601–1623.

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, **10**, 1665–1677.

Xie, M., Kolassa, J., Liu, R., and Liu, D. (2014). Does a zero-total-event study contain information for inference of odds ratio in meta-analysis? Technical report, Rutgers University.

Xie, M., Singh, K., and Strawderman, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, **106**.

Yang, G., Liu, D., and Xie, M. (2012). Tian's exact meta-analysis method as a special example under the general framework of combining cds. Research note.