Biometrical Journal **52** (2010) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

Meta-Analysis of Clinical Trials with Rare Events

Dankmar Böhning[*,1], **Kalliopi Mylona** [1], and **Alan Kimber** [1]

[1] Southampton Statistical Sciences Research Institute & Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, UK

Received zzz, revised zzz, accepted zzz

Meta-analysis of rare event studies has recently become a subject of controversy and debate. We will argue and demonstrate in this paper that the occurrence of zero events in clinical trials or cohort studies, even if zeros occur in both arms (the case of a double-zero trial), is less problematic, at least from a statistical perspective, if the available statistical tools are applied in the appropriate way. In particular, it is neither necessary nor advisable to exclude studies with zero events from the meta–analysis. In terms of statistical tools we will focus here on Mantel-Haenszel techniques, mixed Poisson regression and related regression models.

Key words: Meta-analysis; Rare event analysis; Zero and double-zero studies; Mantel-Haenszel estimation; Poisson regression with random effects;

1 Introduction and background

We are interested in meta–analysis of clinical trials with binary endpoints and with the occurrence of rare events. A *rare event* here means that the event occurrence probability is so low that frequently a small number or no events are observed in a trial, despite the fact that either the trial sizes or the observation times are not small. Hence it is different from meta-analysis of clinical trials with sparse events where trial sizes are small (often for reasons of patient recruitment) but event probabilities might not necessarily be small. Meta-analysis allows the researcher to reach conclusions based on a set of independently performed studies. Provided that the information on an intervention effect is reliable, meta-analysis is a powerful tool, used for analyzing and combining the results obtained from individual studies (Böhning *et al.,* 2008). However, there are potential weaknesses when designing and performing a meta-analysis, which is why controversy often arises among researchers.

The work is motivated by a recent debate on the cardiovascular safety of the diabetes drug *Rosiglitazone* which arose after a publication of a meta–analysis that showed a significantly elevated risk for myocardial infarction (MI) and a borderline significant increased risk for cardiovascular (CV) mortality (Nissen and Wolski 2007). Some meta–analyses confirmed the original findings (Singh *et al.* 2007) whereas others reported inconclusive findings (Diamond *et al.* 2007).

After the publication of the meta-analysis conducted by Nissen and Wolski (2007), numerous scientists carried out their own analyses, in order to assess the 'true' effect of Rosiglitazone with respect to the occurrence of MI and CV deaths, including the meta-analyses of Home *et al.* (2007), Bracken (2007), Diamond *et al.* (2007), Shuster *et al.* (2007), Dahabreh (2008), Tian *et al.* (2009), Friedrich *et al.* (2009), Mannucci *et al.* (2009), and Cai *et al.* (2010).

*Corresponding author: e-mail: d.a.bohning@soton.ac.uk, Phone: +44 23 8059 6712, Fax: +44 23 8059 3216

As been pointed out in Kaul and Diamond (2011), one issue of this debate is whether trials with zero events should be excluded from the analysis and what biasing effect this exclusion would have. We will argue here that, at least from a statistical perspective, *such exclusions can and should be avoided* if the available statistical repertoire is used appropriately.

The standard approach to meta-analysis assumes that an estimate of the effect measure of interest is available from each study together with an estimate of its variance which is then typically treated as a known parameter. In the standard approach, it is further assumed that the effect measure is normally distributed within studies and all further investigations, such as heterogeneity modelling, build on these assumptions. This approach is reviewed in Stijnen *et al.* (2010) and it is pointed out that the approach has several shortcomings when the effect-measure involves count data and therefore is very relevant here. Instead, Stijnen *et al.* (2010) highlight the benefits that occur when exact methods, reflecting the count nature of the data involved in the meta-analysis, are used. They point out that the bias in the standard approach, caused by the correlation between estimate and standard error, can be avoided. As a second benefit, the use of a more appropriate within-study likelihood incorporates the uncertainty in the estimates of the standard errors and, hence, provides a more realistic approach. In addition, it avoids the use of continuity corrections. This point is most relevant for meta-analysis of studies with many single– or double–zero studies where the study-specific effect measure itself, the risk ratio say, would not be estimable without the use of a continuity correction. This is crucial for the standard approach since it builds on study-specific effect measures. Here, very much in the spirit of the approach taken by Stijnen *et al.* (2010), we focus on alternative approaches that avoid the use of continuity corrections.

The paper is organized as follows. In section 3 we will focus on Mantel-Haenszel methodology, which has the property of being robust with respect to the occurrence of zero-events, whereas in section 4 we focus on Poisson modelling. In particular, we demonstrate that the question of homogeneity of effect can be investigated using a random effect for the study factor. In addition, we mention zero–inflation modelling as an option to check whether a large number of zeros in the meta-analytic data would require a zero–inflation component. The paper ends with a short discussion.

2 Data

Recently, Nissen and Wolski (2010) published a second meta-analysis on Rosiglitazone including 56 trials with in total 35531 patients. The inclusion criterion for a trial was a duration of at least 24 weeks. Their current findings suggest an increase in risk ratio for Rosiglitazone, as the Rosiglitazone therapy significantly increased the risk of MI, but not CV mortality. We take this data set as the most complete data basis and all our analysis is grounded on the data reproduced in Tables 1 and 2. Note that the trials differ in observation time, ranging from 24 weeks to 260 weeks. Hence the person-time (= size of trial $\times$ duration) differs across trials and this needs to be taken into account.

Forest plots for MI events and CV events are provided in Figures 1 and 2, respectively, using the package `STATA`. These not only show the distribution of the risk ratio across studies, they also illuminate how many studies have been excluded by `STATA` for the analysis due to zero events.

3 Mantel-Haenszel techniques

Let x_i^T and x_i^C denote the number of events (CV or MI) in the treatment and in the control arms respectively of the $i-$th trial. Further denote by P_i^T and P_i^C the person-time in the treatment and in the control arms respectively of the $i-$th trial. Also let $x^T = \sum_{i=1}^k x_i^T$ denote the total number of events in the k trials for the treatment arm with similar definitions for x^C, P^T and P^C. Then the *crude* risk ratio is simply $\widehat{RR} = \frac{x^T P^C}{x^C P^T}$ which relates the estimated overall risk x^T/P^T in the treatment arm to the estimated overall risk x^C/P^C in the control arm. The calculation of the crude risk ratio is straightforward unless x^T or x^C

Table 1: Study data for the meta–analysis on rare events in the Rosiglitazone and control arms; MI refers to myocardial infarction events, CV to cardiovascular deaths, n is the size and P is the person-time of the respective study arm and 'duration' refers to the study period at risk (in weeks)

		treatment arm				control arm				
ID	study label	n	P	MI	CV	n	P	MI	CV	duration
1	49653/011	176	4224	0	0	357	8568	2	1	24
2	49653/020	207	10764	1	0	391	20332	2	0	52
3	49653/024	185	4810	1	0	774	20124	1	0	26
4	49653/093	109	2834	1	0	213	5538	0	0	26
5	49653/094	116	3016	0	0	232	6032	1	1	26
6	100684	47	2444	1	0	43	2236	0	0	52
7	49653/143	124	2976	0	0	121	2904	1	0	24
8	49653/211	114	5928	2	4	110	5720	5	5	52
9	49653/284	384	9216	0	0	382	9168	1	0	24
10	712753/008	135	6480	0	0	284	13632	1	0	48
11	AVM100264	302	15704	1	1	294	15288	0	2	52
12	BRL49653C/185	142	4544	0	0	563	18016	2	0	32
13	BRL49653C/334	279	14508	1	1	278	14456	2	0	52
14	BRL49653C/337	212	5088	0	0	418	10032	2	0	24
15	49653/015	198	4752	1	0	395	9480	2	2	24
16	49653/079	106	2756	1	1	203	5278	1	1	26
17	49653/080	99	15444	2	0	104	16224	1	0	156
18	49653/082	107	2782	0	0	212	5512	2	1	26
19	49653/085	139	3614	1	0	138	3588	3	1	26
20	49653/095	96	2496	0	0	196	5096	0	1	26
21	49653/097	120	18720	1	0	122	19032	0	0	156
22	49653/125	173	4498	1	0	175	4550	0	0	26
23	49653/127	58	1508	0	0	56	1456	1	0	26
24	49653/128	38	1064	0	0	39	1092	1	0	28
25	49653/134	276	7728	2	0	561	15708	0	1	28
26	49653/135	111	11544	3	1	116	12064	2	2	104
27	49653/136	143	3718	0	0	148	3848	1	2	26
28	49653/145	242	6292	0	0	231	6006	1	1	26
29	49653/147	88	2288	0	0	89	2314	1	0	26
30	49653/162	172	4472	0	0	168	4368	1	1	26

continued in Table 2

is zero, a situation that can be excluded in nearly all practical cases and, in particular, for the meta–analysis at hand.

However, it is important in any meta–analysis to investigate the factor *study* as a potential confounding factor and one way to do this is to stratify. Calculation of a weighted estimator with weights being calculated on the basis of the inverse variance is almost impossible for at least two reasons. First, the calculation of the study specific risk ratio $\widehat{RR}_i = \frac{x_i^T P_i^C}{x_i^C P_i^T}$ is prohibited if zero events occur, and secondly, the variance of $\widehat{RR}_i$ is difficult to compute reasonably accurately and the existing estimator, based on the the $\delta-$method and treating the estimated weights as non-random, becomes an unstable estimator when zero-events occur (Böhning and Sarol 2000). This problem could be addressed by using continuity corrections (adding a constant to all cells, as suggested in Jewell (2004, p. 80)) as in Kaul and Diamond (2011), but

Table 2: Study data for the meta–analysis on rare events in the Rosiglitazone and control arms (continued from Table 1)

		treatment arm				control arm				
ID	study label	n	P	MI	CV	n	P	MI	CV	duration
31	49653/234	61	1586	0	0	116	3016	0	0	26
32	49653/330	377	19604	0	0	1172	60944	1	1	52
33	49653/331	325	16900	0	0	706	36712	0	1	52
34	49653/137	185	5920	2	1	204	6528	1	0	32
35	SB-712753/002	280	6720	0	0	288	6912	1	1	24
36	SB-712753/003	272	8704	0	0	254	8128	1	0	32
37	SB-712753/007	154	4928	0	0	314	10048	1	0	32
38	SB-712753/009	160	3840	0	0	162	3888	0	0	24
39	49653/132	112	2688	0	0	442	10608	1	1	24
40	AVA100193	124	2976	0	0	394	9456	1	1	24
41	AVD102209	131	10218	0	1	132	10296	0	0	78
42	AVD104742	213	5538	0	0	160	4160	0	0	26
43	AVD100521	337	9436	7	3	331	9268	8	4	28
44	AVA105640	250	6500	1	1	331	8606	1	0	26
45	ARA102198	49	1176	0	0	49	1176	0	0	24
46	49653/044	51	1326	0	0	101	2626	0	0	26
47	49653/096	115	2990	0	0	232	6032	0	0	26
48	49653/109	25	650	0	0	52	1352	0	0	26
49	49653/325	195	4680	0	0	196	4704	0	0	24
50	49653/282	75	1800	0	0	70	1680	0	0	24
51	49653/351	29	1508	0	0	28	1456	0	0	52
52	49653/369	24	624	0	0	25	650	0	0	26
53	49653/452	24	576	0	0	26	624	0	0	24
54	DREAM	2634	410904	9	10	2635	411060	15	12	156
55	ADOPT19	2895	602160	41	5	1456	302848	27	2	208
56	RECORD	2227	579020	56	71	2220	577200	64	60	260

these corrections often add bias of unclear size and direction. See also the debate on adding something to nothing in Sweeting *et al.* (2004), Rücker *et al.* (2009), Shuster *et al.* (2007) and Friedrich *et al.* (2009). Furthermore, Bhaumik *et al.* (2012) show that, if a continuity correction is used, a constant value of $\frac{1}{2}$ removes the first-order bias.

Fortunately, the need to estimate study–specific risk ratios is unnecessary with Mantel-Haenszel methods. The beauty of Mantel-Haenszel methods can be seen in the fact that they follow the rule *sums before ratios* which leads to their celebrated robustness properties. The Mantel-Haenszel estimate of relative risk (Clayton and Hills 1993; Jewell 2005) is

$$\widehat{RR}_{\text{MH}} = \frac{\sum_i x_i^T P_i^C / P_i}{\sum_i x_i^C P_i^T / P_i}, \tag{1}$$

where $P_i = P_i^C + P_i^T$. Note that $\widehat{RR}_{\text{MH}}$ is only undefined when x^T or x^C is zero. It also can be viewed as a weighted sum $\frac{\sum_i w_i \widehat{RR}_i}{\sum_i w_i}$ of the $\widehat{RR}_i$'s with weights $w_i = x_i^C P_i^T / P_i$. We point out that the computational form to be actually used is given by (1) since the weighted version would remove any study with at least one arm having zero–events.

Table 3 provides the analysis for MI and CV mortality. We have used the package `STATA` with more details given in the supplementary material. For MI we see a slight confounding (masking) effect making

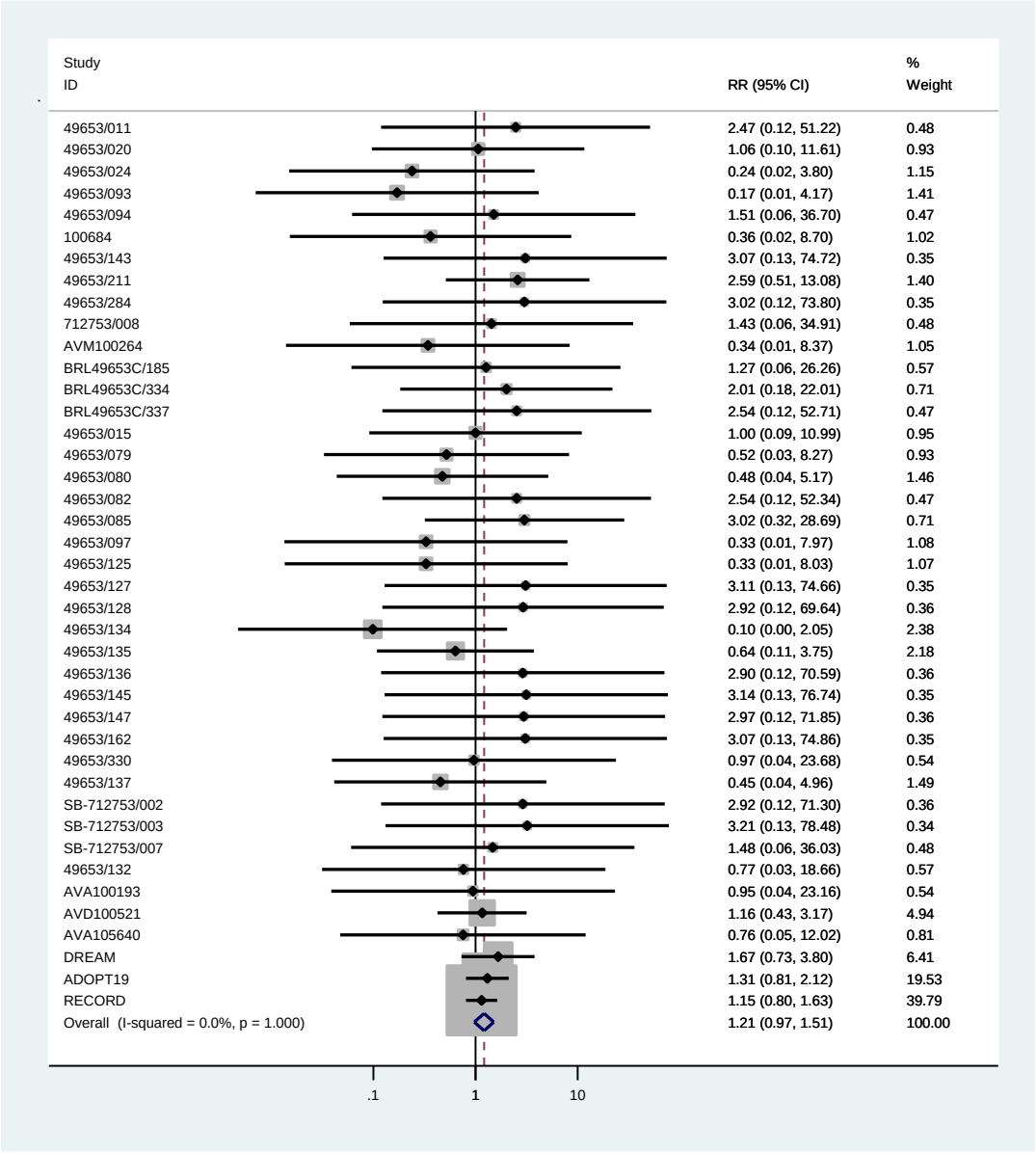


Figure 1: Forest plot for the risk ratio of MI events; 15 double-zero trials had to be excluded for the construction of this plot

the Mantel-Haenszel estimate borderline significant. For CV there is a more pronounced confounding (inflation) effect leading to non-significant Mantel-Haenszel estimate, only slightly above the null-effect.

One of the benefits of using the Mantel-Haenszel (MH) estimate is that we can clearly see the effect of including/excluding single-zero or double-zero trials. Note that zero events will not change the denominator or numerator in the MH estimate. Hence, in the case of a double-zero trial, neither denominator

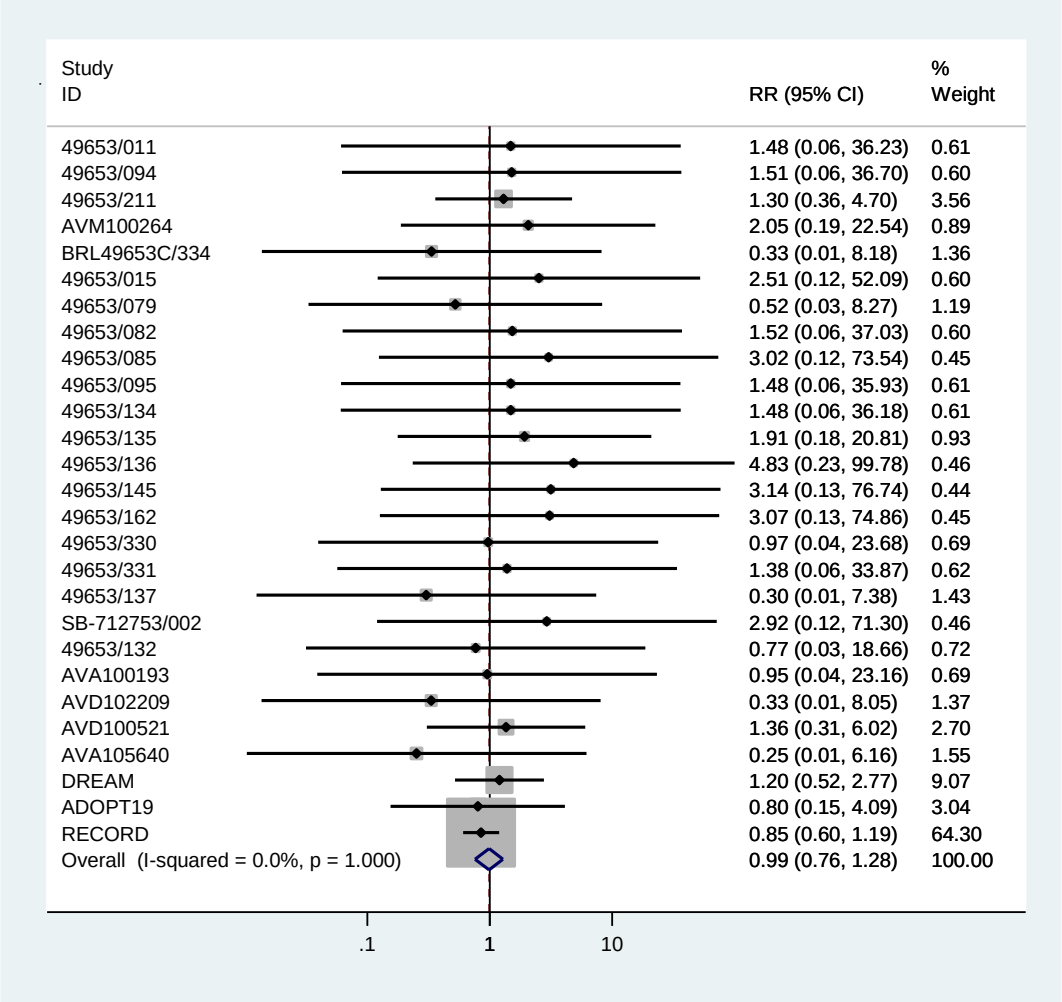


Figure 2: Forest plot for the risk ratio of CV deaths; 29 double-zero trials had to be excluded for the construction of this plot

or numerator will be affected by the exclusion or inclusion of the trial. In the case of a single-zero trial, the numerator or denominator of the MH estimate will be affected depending in which arm the non-zero event occurs. Hence the estimator might experience bias if single-zero trials are excluded. Hence we argue here to include *all* trials, single- and double-zero trials, in the estimation. This is also in line with other evidence, for example in Friedrich *et al.* (2007) or Bhaumik *et al.* (2012). In addition, it might be argued on ethical grounds that patients have the right to have their data stemming from zero trials entered into the meta-analysis (Keus *et al.* 2009). We also investigate in Table 3 the effect of excluding zero-studies from the analysis on the MH estimate. As expected, there is no effect of excluding double-zero studies on the MH estimate of the risk ratio. More serious are the effects when single-zero trials are excluded. In the

Table 3: Mantel-Haenszel (MH) estimate, defined in (1), in the rare events meta–analysis of Rosiglitazone and the crude risk estimate given as $\widehat{RR} = \frac{x^T P^C}{x^C P^T}$; the effect of excluding double-zero (DZ) and single-zero (SZ) studies on the MH estimate; number of studies included is given in brackets in the first column

method	estimate	confidence interval	P-value
MI			
crude(56)	1.2561	0.9928 – 1.5911	0.0504
MH(56)	1.2782	1.0125 – 1.6137	0.0390
DZ(41)	1.2782	1.0125 – 1.6137	0.0390
SZ(15)	1.2097	0.9489 – 1.5422	0.1244
CV			
crude(56)	1.1281	0.8496 – 1.4987	0.4051
MH(56)	1.0257	0.7760 – 1.3557	0.8585
DZ(27)	1.0257	0.7760 – 1.3557	0.8585
SZ(8)	0.9374	0.7015 – 1.2526	0.6620

case of MI-events, the significance of the effect is lost and in the case of CV-events the effect, although not significant, changes from harmful to protective. In summary, in the case of a double-zero trial, the MH-estimate does not change in either direction, which seems to be a desirable property as there is no evidence in a double-zero trial of benefit in either the treatment or control arms. In the case of a single-zero trial, the MH-estimate changes in favor of treatment or control depending where the non-zero event occurs.

The major difficulty in MH estimation with rare events, however, lies in investigating homogeneity of effect. There exists a χ^2-test of homogeneity which, unfortunately, requires both stable study–specific effect estimates and stable study–specific variance estimates of the study–effects. Hence the available χ^2-test of homogeneity is of unknown behaviour even if infeasible study-specific effect estimates are omitted. In the following we present a modelling approach based on random effects which allows the homogeneity of effect to be investigated in a straightforward way.

4 Poisson regression

It was seen in the previous section that MH estimation provides a simple and powerful tool for adjusting the risk ratio for the potentially confounding study factor. In this section we turn to regression models as these can incorporate additional covariates as fixed and/or random effects. This will also allow a more satisfying way of dealing with effect heterogeneity. The major idea of Poisson modelling is to consider the count of events X as a Poisson distributed variable with mean $E(X) = \mu P$ (Breslow and Day 1987; Clayton and Hills 1993). Evidently, $\mu = E(X)/P$ is the incidence rate. We have that for each trial i and each treatment arm j $E(X_{ij}) = \mu_j P_{ij}$ where now $j = 1$ means being on the treatment arm and $j = 0$ otherwise. Hence the risk ratio is $RR = \mu_1/\mu_0$. Taking logarithms yields the basic *log-linear model*:

$$\log E(X_{ij}) = \log P_{ij} + \log \mu_j = \log P_{ij} + \alpha + \beta \times j, \tag{2}$$

where often the notation β for the log-risk ratio is used and α is the baseline risk. Note that the model allows non–identical within–trial person times. This is a slightly more general formulation of the model than necessary for the data set at hand but we leave it this way for the sake of generality. Parameter estimates are found by maximizing the associated Poisson likelihood. Since the basic model does not involve *study* as a factor the maximum likelihood estimate of β corresponds to the crude log-risk ratio.

Model (2) has one peculiarity in that it involves the logarithmic person-times $\log P_i$ as an *offset*. An offset represent covariate information with a fixed parameter of 1 attached to it. Most statistical packages, including `STATA`, the package we use in our analyses, have options to include an offset. More details are given in the appendix.

4.1 Poisson regression and random study effect

Clearly, model (2) suffers from the fact that the factor *study* is ignored. This can be easily modified to include the study effect as

$$\log E(X_{ij}) = \log P_{ij} + \log \mu_{ij} = \log P_{ij} + \alpha_i + \beta_i \times j. \tag{3}$$

Model (3) not only allows different study-specific baseline risks α_i, but also study-specific log-risk ratios β_i. Assuming $\beta_i = \beta$ for all i leads to an estimate of β that is equivalent to the MH analysis. Let $\eta_{ij} = E(X_{ij})$, so that the Poisson likelihood becomes for the common effect model

$$\prod_i \prod_j Po(x_{ij}|\eta_{ij}) = \prod_i \left[Po(x_{i0}|P_{i0}\exp(\alpha_i)) \times Po(x_{i1}|P_{i1}\exp(\alpha_i+\beta))\right], \tag{4}$$

where the $Po(x|\eta)) = \exp(-\eta)\eta^x/x!$ are Poisson probabilities. In this likelihood (4) *study* occurs as a fixed effect. However, it is a common understanding that in this situation *study* should be a random effect. This means that α_i is not considered as a fixed but unknown parameter. Instead it is assumed to be random quantity, here as normal with unknown mean α and unknown variance σ_α^2. In this case, likelihood becomes

$$\prod_i \int \left[Po(x_{i0}|P_{i0}\exp(\alpha_i)) \times Po(x_{i1}|P_{i1}\exp(\alpha_i+\beta))\right] \phi(\alpha_i|\alpha,\sigma_\alpha^2) d\alpha_i, \tag{5}$$

where $\phi(\alpha_i|\alpha,\sigma_\alpha^2)$ denotes the probability density of a normal random variable with mean α and standard deviation σ_α. Note that the order of products in (5) is no longer exchangeable and, hence, this better reflects the split-plot character of the data (*treatment* varies only within *study*). This approach is preferred for the following reasons:

- It avoids the so-called Neyman–Scott problem, meaning that there could arise a consistency problem since the number of parameters in the α_i is connected to the number of studies. Hence the number of parameters will increase with the number of studies. The random effect approach avoids this (only one variance parameter throughout) and it is possible to estimate the random effects distribution consistently (Kiefer and Wolfowitz 1956).
- Finally, considering *study* as a fixed effect could lead to unstable parameter estimates, as is the case here because of the rare event nature of the data.

Another benefit of this approach is that it may easily be generalized to include a random effect for the log-risk ratio, $\beta_i \sim N(\beta, \sigma_\beta^2)$, again a mean β normal distribution with variance σ_β^2. The likelihood then becomes

$$\prod_i \int Po(x_{i0}|P_{i0}\exp(\alpha_i)) \times \left[\int Po(x_{i1}|P_{i1}\exp(\alpha_i+\beta_i))\phi(\beta_i|\beta,\sigma_\beta^2)d\beta_i\right] \phi(\alpha_i|\alpha,\sigma_\alpha^2)d\alpha_i. \tag{6}$$

The integrals are usually approximated by Hermite-Gaussian quadrature (Aitkin 1999) but other techniques including the Laplacian approximation are available as well. For more details on computation with he software `STATA` see the supplementary material. The likelihoods (4), (5), and (6) can be used in likelihood-ratio testing of whether random effects are necessary and whether they should involve a random intercept or a random slope or both forms. We exemplify these ideas in the following.

The results of the analysis are provided in Table 4. For both MI and CV we see that a random study-effect for the intercept is needed, but the random effect for the log relative risk, as estimated by σ_β^2, is virtually zero in both cases. Note that also a more formal evaluation of models is possible using the likelihood ratio test (LRT). If Log-L_1 and Log-L_0 are the maximised log-likelihoods associated with two models, M_1 and

Table 4: Poisson regression estimates in the rare events meta–analysis of Rosiglitazone; Log-L stands for the maximised log-likelihood under the respective model; σ^2_α and σ^2_β refer to the variance of the random intercept and random slope, respectively

Poisson model	estimate	confidence interval	Log-L
MI			
treatment	1.2561	0.9991 – 1.5793	-174.2054
treatment	1.2634	1.0006 – 1.5952	-137.9566
σ^2_α	0.6352	0.3213 – 1.2559	
treatment	1.2634	1.0006 – 1.5952	-137.9566
σ^2_α	0.6352	0.3213 – 1.2559	
σ^2_β	0.		
CV			
treatment	1.1281	0.8579 – 1.4835	-172.0216
treatment	1.0192	0.7737 – 1.3426	-100.3147
σ^2_α	1.2328	0.5908 – 2.5723	
treatment	1.0192	0.7737 – 1.3426	-100.3147
σ^2_α	1.2328	0.5908 – 2.5723	
σ^2_β	0.		

M_0, under consideration, then the test statistic of the likelihood ratio test is given as $2(\text{Log-L}_1 - \text{Log-L}_0)$ assuming that model M_0 is nested in model M_1 (M_0 occurs as a special case of M_1). Note that roles of M_1 and M_0 change when applied to Table 4 as the three models are nested in increasing order. Under M_0, this test statistic is approximately distributed as χ^2 where the degrees of freedom are determined as the difference of the number of parameters involved in M_1 and M_0. As a cautionary note we add that the conventional asymptotic χ^2-result is only valid under the assumption that M_0 is not on the boundary of M_1. However, this assumption is violated when testing $M_0 : \sigma^2_\alpha = 0$ against $M_1 : \sigma^2_\alpha > 0$. In this case, the null-distribution of the likelihood ratio statistic is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ (Self and Liang 1987) where χ^2_0 is the one-point distribution at 0. In practice this means that ordinary p-values have to be divided by 2 to get the correct asymptotic p-values. Applying this test to the log-likelihood values in Table 4, we confirm that the LRT for comparing the random intercept with the fixed intercept model is highly significant, whereas the LRT comparing the random intercept and random slope model with the random intercept-only model is non-significant (the log-likelihoods are virtually identical). We conclude that there is a *homogeneous* treatment effect which is borderline significant for MI but not significant for CV mortality.

In Table 5 we investigate, on the basis of model (5), the effect of excluding zero-studies from the analysis. As expected, the effect of excluding double-zero studies on the risk ratio is rather minor with almost negligible impact on standard errors and confidence intervals, at least as far as MI-events are concerned but with slightly more change for CV-events. More serious are the effects when single-zero trials are excluded. It can be seen in Table 5 that, for MI-events, the borderline significance is lost, with a confidence interval now clearly including the reference value. But also for CV-events, there is an interesting change: although not significant the effect crosses the reference value and becomes protective. Hence great care must be taken when excluding zero trials and and ideally such exclusions should be avoided.

We mention briefly that the Poisson regression model with random intercept and random slope can be extended to allow correlation between these two random effects. The associated likelihood is provided as

$$\prod_i \left[\int \int Po(x_{i0}|P_{i0}\exp(\alpha_i)) \times Po(x_{i1}|P_{i1}\exp(\alpha_i+\beta_i))\phi(\alpha_i,\beta_i|\alpha,\beta,\Sigma) \right] d\beta_i d\alpha_i, \qquad (7)$$

Table 5: Poisson random effects regression estimates of the risk ratio in the rare events meta–analysis of Rosiglitazone: the effect of excluding DZ and SZ studies and none excluded (NONE); number of studies included is given in brackets in the first column

zero-studies (k)	RR-estimate	SE	Z	P-value	95% CI
MI					
NONE(56)	1.2633	0.1503	1.96	0.049	1.0006 – 1.5952
DZ(41)	1.2634	0.1503	1.97	0.049	1.0008 – 1.5955
SZ(15)	1.2101	0.1512	1.53	0.127	0.9473 – 1.5458
CV					
NONE(56)	1.0193	0.1433	0.14	0.892	0.7738 – 1.3426
DZ(27)	1.0246	0.1441	0.17	0.863	0.7778 – 1.3497
SZ(8)	0.9427	0.1395	-0.40	0.690	0.7054 – 1.2599

where $\phi(\alpha_i, \beta_i|\alpha, \beta, \Sigma)$ is the bivariate normal density with mean vector elements α and β, and covariance matrix Σ with elements σ^2_α, σ^2_β on the diagonal and covariance $\sigma_{\alpha,\beta}$. This model can be also fitted in STATA although we have not done so here since already the independence model showed that the random slope effect does not yield any significant increase in the likelihood.

4.2 *Zero-inflation models*

We have seen above that more than 50% of the trials involved in the meta-analysis have zero–events in at least one arm. Dealing with count data in which there are many zero counts leads naturally to the question whether these represent an excess of zero counts relative to the Poisson model. An excess in zero counts is also called *zero–inflation*. Zero-inflated Poisson (ZIP) models have become an accepted methodology to cope with excess zeros. The original work by Lambert (1992) suggests a way of modelling count data with excess zeros as follows. It is assumed that there is a compartment that generates only zero counts and which occurs with probability α. Furthermore, it is assumed that outside this compartment the regular Poisson model holds. This occurs with probability $(1 - \alpha)$, evidently. Hence we have the following ZIP–model, adapted to our situation

$$Pr[X_{ij} = 0] = \quad \pi_{ij} + (1 - \pi_{ij})e^{-\eta_{ij}} \tag{8}$$

$$Pr[X_{ij} = x] = \quad (1 - \pi_{ij})Po(x|\eta_{ij}) \text{ for } x = 1, 2, ... \tag{9}$$

where i is the trial number index and $j = 0, 1$ indicates the trial arm – as before. The Poisson part $Po(x|\eta_{ij})$ of (9) is modelled as previously. A feature of the Lambert model that the excess zero part is modelled by means of a logistic regression approach, which leads to

$$\log \eta_{ij} = \quad \log P_{ij} + \log \mu_{ij} = \log P_{ij} + \quad \alpha + \beta \times j \tag{10}$$

$$\text{logit } \pi_{ij} = \quad \log \pi_{ij} - \log(1 - \pi_{ij}) = \quad \alpha' + \beta' \times j. \tag{11}$$

In this formulation (10) and (11) have the same covariates occurring, though this is not a requirement. Different covariates may occur in (10) and (11). Note that here β is a log-relative risk whereas β' is a log-odds ratio. The ZIP regression models are easy to fit and are available in many packages including STATA. Note that there is no offset term in (11). Although this is technically possible it is rarely meaningful for the logistic part of the ZIP–model. Whereas it is reasonable to assume that the average count of cases is linearly related to the amount of person-time, it is not plausible to assume that the probability of an extra zero count is linearly related to the person-time. STATA offers an offset term for both parts of (10) and (11), so it is important to see that the offset term is only appropriate for (10).

Table 6: Zero-inflated Poisson regression estimates in the rare events meta–analysis of Rosiglitazone; Log-L stands for the maximised log-likelihood

model	estimate	confidence interval	Log-L
MI			
– ZIP model with treatment effect in both logistic and Poisson parts –			-174.1394
		Poisson part	
$\exp(\beta)$	1.2685	1.0034 – 1.6036	
		logistic part	
β'	11.9924	-2739.2 – 2763.2	
– ZIP model with treatment effect only in Poisson part –			-174.1943
		Poisson part	
$\exp(\beta)$	1.2600	0.9987 – 1.5895	
– logistic part has constant zero-inflation with $\hat{\pi} = 0.0060$ –			
– standard (no zero–inflation) Poisson model –			-174.2054
$\exp(\beta)$	1.2561	0.9991 – 1.5793	
CV			
– ZIP model with treatment effect in both logistic and Poisson parts –			-171.7274
		Poisson part	
$\exp(\beta)$	1.0953	0.8257 – 1.4530	
		logistic part	
β'	-16.3474	–12095. – 12063.	
– ZIP model with treatment effect only in Poisson part –			-171.9951
		Poisson part	
$\exp(\beta)$	1.1310	0.8588 – 1.4894	
– logistic part has constant zero-inflation with $\hat{\pi} = 0.0287$ –			
– standard (no zero–inflation) Poisson model –			-172.0216
$\exp(\beta)$	1.1281	0.8579 – 1.4835	

We have applied ZIP modelling to the meta-analytic data at hand and the results are shown in Table 6. The conventional Poisson regression model is compared with the ZIP model where there is constant zero-inflation as well as assuming a treatment effect in the logistic part. Note that we have included the estimate of the proportion of extra-zeros in Table 6 in the case of constant inflation. On the basis of the LRT, neither of the two comparisons is significant for either endpoint. We conclude that, although the data contain many zeros, these are compatible with the conventional Poisson regression model. Finally, we address the question how a random effect could be supplemented to the zero-inflated Poisson regression model. This can be accomplished by allowing α_i to be a normal mean α random-effect with variance σ^2_α, $\alpha_i \sim N(\alpha, \sigma^2_\alpha)$, and, if desired, by allowing $\alpha'_i \sim N(\alpha', \sigma^2_{\alpha'})$:

$$\log \eta_{ij} = \log P_{ij} + \log \mu_{ij} = \log P_{ij} + \alpha_i + \beta \times j \tag{12}$$

$$\text{logit}\, \pi_{ij} = \log \pi_{ij} - \log(1 - \pi_{ij}) = \alpha'_i + \beta'_i \times j, \tag{13}$$

where i is the trial number index and $j = 0, 1$ indicates the trial arm – as before. This is more for completeness than for reasons of necessity in this case, since there is no evidence of any zero-inflation. The model of interest for our case would be the Poisson regression part with random intercept effect supplemented by constant zero-inflation $\pi_{ij} = \pi$ for all studies i and both arms $j = 0, 1$. The associated likelihood is given by

$$\prod_i \int \left\{ \prod_{j=0,1} [\delta_0\{x_{ij}\}\pi + (1-\pi) Po(x_{ij}|P_{i1} \exp(\alpha_i + \beta \times j))] \right\} \phi(\alpha_i|\alpha, \sigma^2_\alpha) d\alpha_i, \tag{14}$$

where $\phi(\alpha_i|\alpha, \sigma_\alpha^2)$ denotes the probability density of a normal random variable with mean α and standard deviation σ_α and $\delta_0\{x_{ij}\} = 1$ if $x_{ij} = 0$ and 0 otherwise. Unfortunately, there is no computational way that this model can be fitted in `STATA`. However, it is possible to fit the models (12) and (13) with *proc nlmixed* in SAS (SAS 2008). We provide details in the appendix. For our case, we only consider testing the Poisson part of the ZIP-model with constant inflation against the ZIP model with constant inflation where the Poisson part is supplemented by a random intercept effect. Not surprisingly, the likelihood ratio test is not significant, neither for MI-events nor for CV-events; in fact, both likelihoods are virtually identical. In the appendix, we also provide the associated versions of `proc nlmixed` which would fit models (12) and (13) simultaneously.

5 Discussion

We have seen that MH techniques and Poisson regression with random effects lead to almost identical results. The benefit of using a Poisson model approach lies in its ability to include additional covariates although in our case no further covariates were available. Whilst we are interested primarily in the treatment effect, the Poisson model also captures variation in the baseline event risks which is most visible in the random effects intercept variance σ_α^2, a term that showed up significantly for both, MI and CV events. Hence there is considerable baseline risk variation across trials.

Another question relates to the issue of interpretation of the observed risk ratio estimate of 1.27 for MI events. We recall that the risk ratio is only one effect measure among others, but a very popular one. Also, it is a relative measure as it ignores the baseline risk. As one may say, a risk ratio of 100 leads also to zero cases if the baseline risk is negligible. For the choice of outcome measure, see also Arends *et al.* (2003) for more. In contrast, absolute effect measures such as the risk difference incorporate the size of the baseline risk and neither SZ nor DZ studies create any problems in the estimation process. A large number of DZ studies would only put a lot of weight on the no-effect. Figure 3 shows bubble plots for CV death and MI event rates. Note that the three larger studies have bubbles centered above zero for MI events, whereas they balance on zero for CV deaths. In addition, the risk difference might be used to arrive at more interpretable numbers of cases to be expected under the given scenario. For example, we find a risk difference estimate for MI events of 0.0019 (using now a risk estimate as number of cases divided by number at risk) which corresponds to a number-needed-to-treat (NNT) of 531. This means that, on average, 531 persons need to be treated to have one additional case. This seems to be a considerable number, depending on how widespread the use of the drug is. In any case, all previous meta-analyses on this issue concentrated on relative risk as we have done here.

Clearly, we find it important to include *duration* of the studies as an important component in the modelling, as the longer the follow-up of the study, the more events can be expected. This has been widely ignored in previous analyses. This leads naturally to the concept of incidence rate, sometimes also known as incidence density, x_i^T/P_i^T and x_i^C/P_i^C in trial i for the treatment and control arm, respectively. Often the emphasis has been on odds–ratio analysis (Tian *et al.* 2009). For the situation here we find the concept of incidence rate more appropriate since it accounts for different trial duration. Odds ratio modelling might be more appropriate in situations of meta-analysis where different study types occur, such as case-control studies as well as cohort studies or clinical trials. Here the odds ratio as effect measure would be more appropriate as it can be estimated on the basis of all these study designs. Odds ratio modelling leads naturally to logistic regression

$$\text{logit}\, p_{ij} = \alpha_i + \beta \times j,$$

where p_{ij} is the probability for an event in study i in the $j-$th treatment arm. Also, $\alpha_i = \alpha$ might be a fixed effect or a random effect $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$ as before. Ignoring the varying duration time and only considering the frequency of events and no-events we find an estimated odds ratio of 1.0088 with 95% confidence interval 0.7633 – 1.3332 for CV-events and an odds ratio of 1.2538 with 95% confidence

interval 0.9904 – 1.5872. These results are very close to the results from the Poisson random effects modelling in Table 4. An explanation for this similarity might be the highly balanced nature of the trials in the sense that all trial arms have the same duration (although duration varies across trials). Logistic regression models with random effects can be easily fitted with `STATA`.

Another, elegant way of involving logistic regression is mentioned in Stijnen *et al.* (2010). The basic idea is to consider X_{i1} conditional on $X_i = X_{i1} + X_{i0}$. Since $E(X_{i1}) = \mu_1 P_{i1}$ and $E(X_{i0}) = \mu_0 P_{i0}$, X_{i1} is binomial with size parameter X_i and event parameter

$$q_i = \frac{\mu_1 P_{i1}}{\mu_1 P_{i1} + \mu_0 P_{i0}} = \frac{RR_i \frac{P_{i1}}{P_{i0}}}{RR_i \frac{P_{i1}}{P_{i0}} + 1}.$$

This is remarkable for two reasons. For one, the event parameter involves only the parameter of interest RR_i. Furthermore, notice that its functional form makes it really prone to logistic regression. Indeed,

$$\text{logit}\, q_i = \log \frac{q_i}{1 - q_i} = \log RR_i + \log \frac{P_{i1}}{P_{i0}} = \alpha_i + \log \frac{P_{i1}}{P_{i0}}$$

where the RHS of the above equation can be used for further modelling such as $\alpha_i = \alpha$ (a common risk ratio across studies) or $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$ (a random effect for the risk ratio). Note this model does not involve a treatment effect as we are used to with the models above but rather the risk ratio estimation and modelling works on the intercept in this case. A disadvantage of the approach is that it has to exclude all double–zero studies. Again, these models can easily be fitted with `STATA`.

Returning now to the Poisson regression model, we have seen that another benefit of the Poisson regression model is that it may easily be extended to allow for zero-inflation and the fundamental model has been suggested by Lambert (1992). The basic ZIP model of Lambert adjusts for overdispersion that arises solely from the occurrence of extra–zeros. Generalizations have been made more recently to account for residual overdispersion stemming from the non-zero part of the count distribution. One of these is the zero-inflated negative binomial model which has received much attention (Hilbe 2011). Another generalization is to replace in the logistic part the logistic link-function by other appropriate links such as the probit-link (offered also by `STATA`). In principle, any other cumulative distribution function may be applied here to specify the link function. ZIP models are easy to interpret and they can lead to more refined data analysis. More on zero-inflation models can be found in Lambert (1992), Cameron and Trivedi (1998), Winkelmann (2003) or Zelterman (2006).

Acknowledgements The authors are grateful for all comments received in the editorial process.

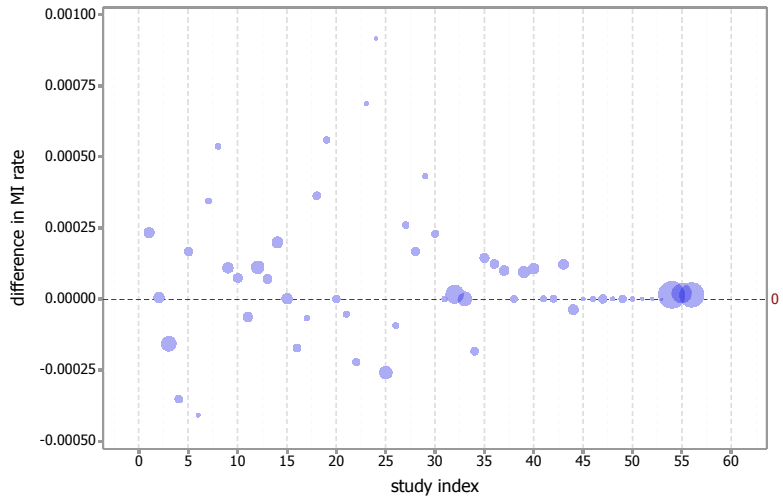
Conflict of Interest
The authors have declared no conflict of interest.

References

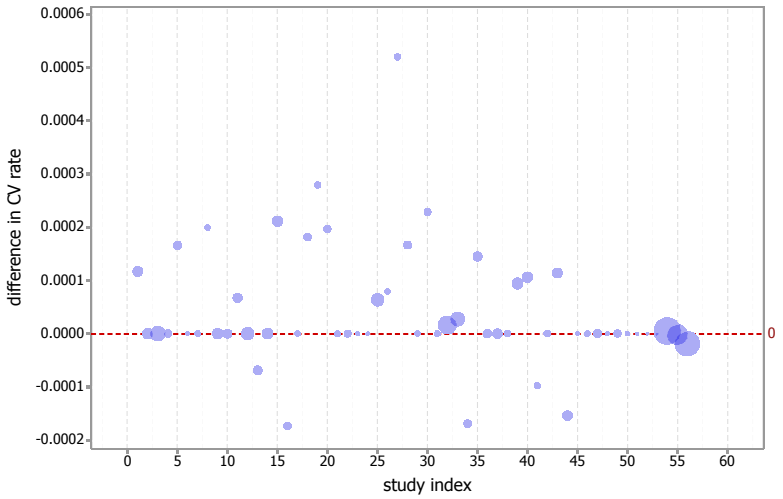
Arends, L.R., Vokó, Z., Stijnen, T. (2003). Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* **22**, 1335-1353.

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

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



(a) event is MI



(b) event is CV

Figure 3: Bubble plot for the risk differences in the rates for MI events and CV deaths; bubble area is proportional to size of study (some smaller bubbles are not perfectly ball-shaped due to the discreteness of the approximation); the study index refers to the ID given in Tables 1 and 2 and the order of studies is the same as in Figures 1 and 2 although DZ studies were excluded in the latter two

Böhning, D. and Sarol, J.Jr. (2000). Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics* **56**, 304–308.

Böhning, D. Kuhnert, R. and Rattanasiri, S. (2008). *Meta-Analysis of Binary Data Using Profile Likelihood.* London, Chapman & Hall.

Bracken, M.B. (2007). Rosiglitazone and cardiovascular risk. *The New England Journal of Medicine* **357**,

937–38.

Breslow, N.E. and Day, N.E. (1987). *Statistical Methods in Cancer Research. Volume 11 - The Design and Analysis of Cohort Studies,* International Agency for Research on Cancer, Lyon.

Cai, T., Parast, L. and Ryan, L. (2010). Meta-analysis for rare events. *Statistics in Medicine* **29**, 2078–2089.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data.* Cambridge, Cambridge University Press.

Clayton, D. and Hills, M. (1993). *Statistical Models in Epidemiology.* Oxford University Press, Oxford.

Dahabreh, I.J. (2008). Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of Rosiglitazone. *Clinical Trials* **5**, 116–120.

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

Friedrich, J.O., Neill, K.J. A., 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**:5.

Friedrich, J.O., Beyene, J. and Adhikari, N.K.J. (2009). Rosiglitazone: can meta-analysis accurately estimate excess cardiovascular risk given the available data? Re-analysis of randomized trials using various methodological approaches. *BMC Research Notes* **2**:5.

Hilbe, J.M. (2011). *Negative Binomial Regression, 2nd Edition.* Cambridge, Cambridge University Press.

Home, P.D., Pocock, S.J., Beck-Nielsen, H., Gomis, R., Hanefeld, M., Jones, N.P., Komajda, M., McMurray, J.J.; RECORD Study Group (2007). Rosiglitazone evaluated for cardiovascular outcomes–an interim analysis. *The New England Journal of Medicine* **357**, 28–38.

Jewell, N.P. (2005). *Statistics for Epidemiology*. Boca Raton, Chapman and Hall/CRC.

Kaul, S. and Diamond, G.A. (2011). Rosiglitazone and cardiovascular risk: Nissen and Wolski 2010 updated meta-analysis revisited. *Journal of the American College of Cardiology* **57**, E552–E552.

Keus, F., Wetterslev, J.,Gluud, C., Gooszen, H.G., van Laarhoven, C.J.H.M. (2009). Robustness assessments are needed to reduce bias in meta-analyses that include zero-event randomized trials. *The American Journal of Gastroenterology* **104**, 546-551.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics* **27**, 887–906.

Lambert, D., (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

Mannucci, E., Monami, M., Di Bari, M., Lamanna, C., Gori, F., Gensini, G.F. and Marchionni, N. (2009). Cardiac safety profile of Rosiglitazone: A comprehensive meta-analysis of randomized clinical trials. *International Journal of Cardiology* **143**, 135–140.

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

Nissen, S.E. and Wolski, K. (2010). Rosiglitazone revisited: an updated meta-analysis of risk for myorcial infaction and cardiovascular mortality. *Archives of Internal Medicine* **170**, 1191–1201.

Rücker, G., Schwarzer, G., Carpenter, J., 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.

SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User's Guide, Second Edition.* Cary, NC, USA, SAS Insitute Inc.

Self, S.G., Liang, K.-Y. (1987). Large sample properties of the maximum likelihood estimator and the likelihood ratio test on the boundary of the parameter space. *Journal of the American Statistical Association* **82**, 605–611.

Shuster, J.J, Jones, L.S., and Salmon, D.A. (2007). Fixed *vs* random effects meta-analysis in rare event studies: The Rosigliazone link with myocardial infarction and cardiac death. *Statistics in Medicine* **26**, 4375–4385.

Singh, S., Loke, Y.K. and Furberg, C.D. (2007). Long-term risk of cardiovascular events with Rosiglitazone: a meta-analysis. *The Journal of the American Medical Association* **298**, 1189–1195.

Stijnen, T., Hamza, T.H., Ozdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* **29**, 3046–3067.

Sweeting, M.J., Sutton, A.J., 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.Y. and Wei, L.J. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2x2 tables with all available data but without artificial continuity correction. *Biostatistics* **10**, 275–281.

Winkelmann, R. (2003). *Econometric Analysis of Count Data.* Springer, New York.

Zelterman, D. (2006). *Models for Discrete Data, revised edition.* Oxford University Press, Oxford.

