

NEJM Rosiglitazone Meta-analysis by Nissan and Wolski - A CSI Investigation

Chuan Zhou
Department of Biostatistics
Vanderbilt University

What reported here is just a sensitivity analysis of the meta-analysis reported in the recent NEJM paper. This should not be regarded as a dispute of their analysis, rather it is intended to be complementary.

1 The case

Rosiglitazone is a drug widely used to treat type 2 diabetes mellitus, but its effect on cardiovascular morbidity and mortality has not been determined. In a recent paper [2] published in the New England Journal of Medicine June 14th issue, Dr Nissen from Cleveland Clinic and his colleague conducted a meta-analysis on 42 randomized-controlled clinical trials and found that Rosiglitazone was associated with a significant increase in the risk of myocardial infarction (OR=1.43; 95% CI=(1.03, 1.98); P=0.03) and with a borderline significant increased risk of death from cardiovascular causes(OR=1.64; 95% CI=(0.98, 2.74); P=0.06).

This paper came to my attention as one day my wife, who is a pharmacist, told me that one of her patients took himself off the drug after hearing about this study on the public radio. And I thought “whoa, here is a paper published in a prestigious journal, accompanied by an editorial, and has been causing a lot of buzz in the media, this must be a great study!” So I dug it out and started reading.

The key findings of the paper are listed in their Table 4, which I replicate here so that we can refer to it more easily.

Table 1: Rates of Myocardial Infarction and Death from Cardiovascular Causes

Study	Rosiglitazone Group <i>no. of events/total no. (%)</i>	Control Group <i>no. of events/total no. (%)</i>	Odds Ratio(95% CI)	P Value
Myocardial infarction				
Small trials combined	44/10,285 (0.43)	22/6106 (0.36)	1.45 (0.88-2.39)	0.15
DREAM	15/2,635 (0.57)	9/2,634 (0.34)	1.65 (0.74-3.68)	0.22
ADOPT	27/1,456 (1.85)	41/2,895 (1.42)	1.33 (0.80-2.21)	0.27
Overall			1.43 (1.03-1.98)	0.03
Death from cardiovascular causes				
Small trials combined	25/6,845 (0.36)	7/3,980 (0.18)	2.40 (1.17-4.91)	0.02
DREAM	12/2,635 (0.46)	10/2,634 (0.38)	1.20(0.52-2.78)	0.67
ADOPT	2/1,456 (0.14)	5/2,895 (0.17)	0.80 (0.17-3.86)	0.78
Overall			1.64 (0.98-2.74)	0.06

All except one clinical trials included were intended to study the efficacy of Rosiglitazone on treating type 2 diabetes, not myocardial infarction or cardiovascular related death, therefore these data may not be appropriate for the intended inference. That should be answered more appropriately by long term and large scale safety studies. But here I will not comment on the pharmacology or pharmokinetics of the drug or why it may relate to increase cardiovascular risk, as a biostatistician I would rather focus on the data analysis.

The authors noticed that myocardial infarction events and cardiovascular deaths were rare in many trials, so the odds ratios and 95% confidence intervals were calculated with the use of Peto method [4]. Following the same

argument used in that paper, the Cochran’s Q statistic for testing heterogeneity across studies is not appropriate either as it is essentially an asymptotic Chi-square test depending on large sample size, for small event sizes as for these trials, it conceivably could have very low power. Testing heterogeneity across trials with sparse data is still a challenging problem. This could have serious consequence as there could be large difference between results from random effect models and fixed effect models.

2 Reconstruct the scene

When I looked at Table 4 in [2], my first reaction was “there was no significant association from all the small trials combined, no associations from DREAM and ADOPT trials either, then why suddenly there was significance when all of them combined?” In the following I will re-analyze the data to see if I can replicate these finding. All analyses were done using the following versions of R [1], the operating system, and add-on packages.

- R version 2.5.0 (2007-04-23), i486-pc-linux-gnu
- Locale: en_US.UTF-8
- Base packages: base, datasets, graphics, grDevices, grid, methods, stats, tools, utils
- Other packages: coda 0.11-3, lattice 0.16-2, meta 0.8-2, R2WinBUGS 2.1-4, rmeta 2.14

Let us first read in the data and check the numbers to see if they agree with the data reported in Table 3 in [2].

```
> rosig <- read.csv("rosigitazone.csv", header = T, sep = ",")
> rosig[, 1:7]
```

	study	n.t	e.mi.t	e.death.t	n.c	e.mi.c	e.death.c
1	49653/011	357	2	1	176	0	0
2	49653/020	391	2	0	207	1	0
3	49653/024	774	1	0	185	1	0
4	49653/093	213	0	0	109	1	0
5	49653/094	232	1	1	116	0	0
6	100684	43	0	0	47	1	0
7	49653/143	121	1	0	124	0	0
8	49653/211	110	5	3	114	2	2
9	49653/284	382	1	0	384	0	0
10	712753/008	284	1	0	135	0	0
11	AVM100264	294	0	2	302	1	1
12	BRL49653C/185	563	2	0	142	0	0
13	BRL49653C/334	278	2	0	279	1	1
14	BRL49653/347	418	2	0	212	0	0
15	49653/015	395	2	2	198	1	0
16	49653/079	203	1	1	106	1	1
17	49653/080	104	1	0	99	2	0
18	49653/082	212	2	1	107	0	0
19	49653/085	138	3	1	139	1	0
20	49653/095	196	0	1	96	0	0
21	49653/097	122	0	0	120	1	0
22	49653/125	175	0	0	173	1	0
23	49653/127	56	1	0	58	0	0
24	49653/128	39	1	0	38	0	0
25	49653/134	561	0	1	276	2	0
26	49653/135	116	2	2	111	3	1
27	49653/136	148	1	2	143	0	0
28	49653/145	231	1	1	242	0	0
29	49653/147	89	1	0	88	0	0
30	49653/162	168	1	1	172	0	0

31	49653/234	116	0	0	61	0	0
32	49653/330	1172	1	1	377	0	0
33	49653/331	706	0	1	325	0	0
34	49653/137	204	1	0	185	2	1
35	SB-712753/002	288	1	1	280	0	0
36	SB-712753/003	254	1	0	272	0	0
37	SB-712753/007	314	1	0	154	0	0
38	SB-712753/009	162	0	0	160	0	0
39	49653/132	442	1	1	112	0	0
40	AVA100193	394	1	1	124	0	0
41	DREAM	2635	15	12	2634	9	10
42	ADOPT	1456	27	2	2895	41	5

OK, data look good. Let's apply the Peto method as the authors did. I am using the *metabin* function in the **meta** R package. Note because it is Peto OR method, pooled estimate of OR is well defined regardless of zero cells. But keep it in mind a valid estimate does not necessarily mean a good estimate.

```
> meta.a <- metabin(e.mi.t, n.t, e.mi.c, n.c, studlab = study,
+ data = rosig, method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)
> summary(meta.a)
```

Number of trials combined: 42

	OR	95%-CI	z	p.value
Fixed effects model	1.4283	[1.0309; 1.9788]	2.1431	0.0321
Random effects model	1.4283	[1.0309; 1.9788]	2.1431	0.0321

Quantifying heterogeneity:

tau² = 0; I² = 0% [0%; 9.9%]

Test of heterogeneity:

Q	d.f.	p.value
29.36	41	0.9125

Method: Peto method

Bingo! Q statistic rejects heterogeneity. The overall OR is 1.43, 95% CI is (1.03, 1.98), P value is 0.03. We have duplicated the results for myocardial infarction, so their meta-analysis withheld our scrutiny, right? Well, not so fast. If we look at the ORs and other estimates for individual studies ...

```
> meta.a
```

	OR	95%-CI	%W(fixed)	%W(random)
49653/011	4.4629	[0.2337; 85.2436]	1.22	1.22
49653/020	1.0585	[0.0977; 11.4661]	1.87	1.87
49653/024	0.1388	[0.0041; 4.6633]	0.86	0.86
49653/093	0.0521	[0.0008; 3.2798]	0.62	0.62
49653/094	4.4817	[0.0701; 286.4922]	0.61	0.61
100684	0.1474	[0.0029; 7.4552]	0.69	0.69
49653/143	7.5745	[0.1503; 381.8452]	0.69	0.69
49653/211	2.5039	[0.5575; 11.2460]	4.71	4.71
49653/284	7.4278	[0.1474; 374.3447]	0.69	0.69
712753/008	4.3726	[0.0660; 289.8681]	0.60	0.60
AVM100264	0.1390	[0.0028; 7.0060]	0.69	0.69
BRL49653C/185	3.5043	[0.1104; 111.2813]	0.89	0.89
BRL49653C/334	1.9595	[0.2030; 18.9146]	2.07	2.07
BRL49653/347	4.5248	[0.2403; 85.1955]	1.23	1.23

49653/015	1.0025	[0.0906; 11.0900]	1.84	1.84
49653/079	0.4972	[0.0267; 9.2572]	1.24	1.24
49653/080	0.4850	[0.0499; 4.7191]	2.05	2.05
49653/082	4.5243	[0.2392; 85.5814]	1.23	1.23
49653/085	2.7685	[0.3858; 19.8662]	2.74	2.74
49653/095	NaN	[NaN; NaN]	0.00	0.00
49653/097	0.1331	[0.0026; 6.7086]	0.69	0.69
49653/125	0.1338	[0.0027; 6.7425]	0.69	0.69
49653/127	7.6577	[0.1519; 386.1576]	0.69	0.69
49653/128	7.2020	[0.1429; 363.0780]	0.69	0.69
49653/134	0.0480	[0.0025; 0.9171]	1.22	1.22
49653/135	0.6362	[0.1084; 3.7319]	3.40	3.40
49653/136	7.1436	[0.1417; 360.2229]	0.69	0.69
49653/145	7.7494	[0.1536; 390.9609]	0.69	0.69
49653/147	7.3065	[0.1450; 368.2472]	0.69	0.69
49653/162	7.5671	[0.1501; 381.4611]	0.69	0.69
49653/234	NaN	[NaN; NaN]	0.00	0.00
49653/330	3.7497	[0.0389; 361.0549]	0.51	0.51
49653/331	NaN	[NaN; NaN]	0.00	0.00
49653/137	0.4629	[0.0477; 4.4888]	2.06	2.06
SB-712753/002	7.1866	[0.1425; 362.3241]	0.69	0.69
SB-712753/003	7.9317	[0.1570; 400.6512]	0.69	0.69
SB-712753/007	4.4391	[0.0685; 287.6430]	0.61	0.61
SB-712753/009	NaN	[NaN; NaN]	0.00	0.00
49653/132	3.5022	[0.0266; 461.0907]	0.45	0.45
AVA100193	3.7237	[0.0377; 367.9541]	0.50	0.50
DREAM	1.6517	[0.7407; 3.6830]	16.53	16.53
ADOPT	1.3294	[0.8002; 2.2086]	41.25	41.25

Number of trials combined: 42

	OR	95%-CI	z	p.value
Fixed effects model	1.4283	[1.0309; 1.9788]	2.1431	0.0321
Random effects model	1.4283	[1.0309; 1.9788]	2.1431	0.0321

Quantifying heterogeneity:

$\tau^2 = 0$; $I^2 = 1$ [1; 1.05]; $I^2 = 0\%$ [0%; 9.9%]

Test of heterogeneity:

Q	d.f.	p.value
29.36	41	0.9125

Method: Peto method

What do we see? We see 4 studies were excluded due to no events in either arms. We see odds ratios in 4's, 7's, 0.1's, huge effect sizes. These ORs correspond to 6 trials with crude OR=0, and 20 trials with crude OR=infinity. The Peto's method essentially pooled over these somehow corrected values, even though these individual estimates are based on very few events in one arm and no events in another arm. At this point, I start wondering how reliable Peto method is. For example, how much should I trust the OR=7.57 for a trial that had 1/121 event rate in one arm, and 0/124 in another arm (Study 49653/143)?

3 Change your perspective - guilty by association

So I run a fixed effect model again, but this time using Mantel-Haenszel method.

```
> meta.b <- metabin(e.mi.t, n.t, e.mi.c, n.c, studlab = study,
+ data = rosig, method = "MH", sm = "OR", incr = 0, allstudies = TRUE)
> meta.b
```

	OR	95%-CI	%W(fixed)	%W(random)
49653/011	Inf	[NaN; Inf]	0.00	0.00
49653/020	1.0591	[0.0955; 11.7498]	2.16	2.27
49653/024	0.2380	[0.0148; 3.8233]	2.67	1.71
49653/093	0.0000	[0.0000; NaN]	2.19	0.00
49653/094	Inf	[NaN; Inf]	0.00	0.00
100684	0.0000	[0.0000; NaN]	1.58	0.00
49653/143	Inf	[NaN; Inf]	0.00	0.00
49653/211	2.6667	[0.5064; 14.0430]	3.11	4.76
49653/284	Inf	[NaN; Inf]	0.00	0.00
712753/008	Inf	[NaN; Inf]	0.00	0.00
AVM100264	0.0000	[0.0000; NaN]	1.63	0.00
BRL49653C/185	Inf	[NaN; Inf]	0.00	0.00
BRL49653C/334	2.0145	[0.1816; 22.3448]	1.64	2.27
BRL49653/347	Inf	[NaN; Inf]	0.00	0.00
49653/015	1.0025	[0.0904; 11.1238]	2.20	2.27
49653/079	0.5198	[0.0322; 8.3940]	2.17	1.70
49653/080	0.4709	[0.0420; 5.2765]	3.36	2.25
49653/082	Inf	[NaN; Inf]	0.00	0.00
49653/085	3.0667	[0.3151; 29.8494]	1.61	2.54
49653/095	NaN	[NaN; NaN]	0.00	0.00
49653/097	0.0000	[0.0000; NaN]	1.67	0.00
49653/125	0.0000	[0.0000; NaN]	1.67	0.00
49653/127	Inf	[NaN; Inf]	0.00	0.00
49653/128	Inf	[NaN; Inf]	0.00	0.00
49653/134	0.0000	[0.0000; NaN]	4.44	0.00
49653/135	0.6316	[0.1035; 3.8532]	4.99	4.02
49653/136	Inf	[NaN; Inf]	0.00	0.00
49653/145	Inf	[NaN; Inf]	0.00	0.00
49653/147	Inf	[NaN; Inf]	0.00	0.00
49653/162	Inf	[NaN; Inf]	0.00	0.00
49653/234	NaN	[NaN; NaN]	0.00	0.00
49653/330	Inf	[NaN; Inf]	0.00	0.00
49653/331	NaN	[NaN; NaN]	0.00	0.00
49653/137	0.4507	[0.0405; 5.0123]	3.46	2.27
SB-712753/002	Inf	[NaN; Inf]	0.00	0.00
SB-712753/003	Inf	[NaN; Inf]	0.00	0.00
SB-712753/007	Inf	[NaN; Inf]	0.00	0.00
SB-712753/009	NaN	[NaN; NaN]	0.00	0.00
49653/132	Inf	[NaN; Inf]	0.00	0.00
AVA100193	Inf	[NaN; Inf]	0.00	0.00
DREAM	1.6698	[0.7295; 3.8225]	14.83	19.17
ADOPT	1.3152	[0.8058; 2.1467]	44.62	54.78

Number of trials combined: 42

	OR	95%-CI	z	p.value
Fixed effects model	1.4269	[1.0294; 1.978]	2.1337	0.0329
Random effects model	NaN	[NaN; NaN]	NaN	--

Quantifying heterogeneity:

tau² = 0; H = 1 [1; 1]; I² = 0% [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
6.02	41	1

Method: Mantel-Haenszel method

The results from fixed effects model using Mantel-Haenszel method are almost identical to those from Peto method. But it is easier to see why MH method might give seriously biased estimate for sparse data. The pooled estimate of OR using MH method is calculated by

$$\bar{T}_{MH(OT)} = \frac{\sum_{i=1}^k a_i d_i / n_i}{\sum_{i=1}^k b_i c_i / n_i}, \quad (1)$$

where a_i , b_i , c_i and d_i are the four cells of the 2×2 table for the $i = 1, \dots, k$ studies, n_i is the total number of patients in i th study.

There are 20 trials with non-zero number of events ($a_i \neq 0$) in the treatment arms, and zero events ($c_i = 0$) in the control arms. For each of these 20 studies, it contributes a positive value to the numerator in Equation 1, but 0 to the denominator, and meanwhile the risk differences are small. As a contrast, there are only 6 studies with zero event in the treatment arms, and non-zero number of events in the control arms, which can add more weights to the denominator in Equation 1. For example, for study 49653/011, the data can be represented in the following 2×2 table

	MI	No MI
Rosiligtazone	2(a)	355(b)
Control	0(c)	176(d)

As you can see, these data contribute $2 \times 176 / 533 = 0.66$ to the numerator in Equation 1, but 0 to the denominator. For study 100684, the data are

0(a)	43(b)
1(c)	47(d)

These data contribute 0 to the numerator in Equation 1, and $1 \times 43 / 91 = 0.47$ to the denominator. But there are 14 more trials putting some positive values to the numerator and nothing to the denominator of the pooled OR! The consequence of this imbalance? A biased pooled estimate of odds ratio that suffers from over-estimation!

MH method and Peto method gave the same pooled estimates of OR. But the estimate based on MH method is biased upward. What does it mean? Peto method over-estimates the OR too!

Well, in the above I took a guilty-by-association approach to explain the cause of bias in both MH and Peto method. But if you really want to see a direct explanation on Peto method and you are brave enough to face more formulas, I will give it try.

The pooled estimate of the odds ratio by Peto method is given by

$$\bar{T}_{PETO(OR)} = \exp\left[\frac{\sum_{i=1}^k (a_i - e_i)}{\sum_{i=1}^k v_i}\right], \quad (2)$$

where $e_i = (a_i + b_i)(a_i + c_i) / n_i$ is the expected number of events in the treatment arm under the null hypothesis, $v_i = \frac{(a_i + b_i)(a_i + c_i)}{n_i} \cdot \frac{(c_i + d_i)}{n_i} \cdot \frac{(b_i + d_i)}{n_i - 1}$ is the variance of the difference $a_i - e_i$.

In Equation 2, the denominator is always positive. In the numerator, $a_i - e_i$ can be expressed as $\frac{a_i d_i - b_i c_i}{n_i}$. If $a_i > 0$ and $c_i = 0$, ie., zero event in the control arm but non-zero number of events in the treatment arm, then $a_i - e_i = a_i d_i / n_i > 0$. If $a_i = 0$ and $c_i > 0$, ie., zero event in the treatment arm but non-zero number of events in the control arm, then $a_i - e_i = -b_i c_i / n_i < 0$. Therefore trials with $c_i = 0$ contribute more positive values to the numerator in Equation 2, and lead to over-estimation of the overall odds ratio.

4 Some lab works

Well, the choice of Peto method by the authors was based on the simulation results reported in [4]. But a different set of simulations [5] arrived at different conclusions. In fact, Sweeting et al. [5] suggested a sensitivity analysis using several methods and continuity corrections should be routinely used. An important scenario was missing in Bradburn et al. [4] is when large imbalance exists between the trials with zero event in different arms. In their simulation, the simple Binomial setup also failed to fully capture the occurrence of zeros in the data.

The binomial distribution assumes a monotone relationship between the probability of zero and the sample size. Under binomial model $y \sim \text{Bin}(n, p)$, the probability of zero $\Pr(y = 0) = (1 - p)^n$, so the larger the sample size the smaller the probability of zero. So if the binomial assumption is true, we would expect to see smaller studies have more zero event, but as we will show later, this is not the case with these data. In fact the occurrence of zero doesn't depend on the size of the trials. Another important aspect of the data that the simulation scheme fails to address is the ratio of number of trials with zero event between treatment and control arms. In the following we first show a zero-inflated Binomial model should be used to capture the data. Second, we will show Peto method doesn't perform well under such a condition.

The following results shows that the occurrence of zeros is not significantly associated with sample size, different from what a Binomial model would suggest.

```
> indx <- c(1:42)
> ind.1 <- rosig$e.mi.t + rosig$e.mi.c
> ind.2 <- rosig$e.mi.t * rosig$e.mi.c
> n.c <- rosig$n.c[indx < 41 & ind.1 > 0]
> y.c <- rosig$e.mi.c[indx < 41 & ind.1 > 0]
> pi.c <- sum(y.c)/sum(n.c)
> n.t <- rosig$n.t[indx < 41 & ind.1 > 0]
> y.t <- rosig$e.mi.t[indx < 41 & ind.1 > 0]
> pi.t <- sum(y.t)/sum(n.t)
> tapply(n.c, y.c == 0, summary)

$`FALSE`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  47.0  110.5   156.0   165.6  200.2   302.0

$`TRUE`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  38.0  115.0   142.5   172.8  219.5   384.0

> tapply(n.t, y.t == 0, summary)

$`FALSE`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  39.0  140.5   243.0   295.9  388.8  1172.0

$`TRUE`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  43.0  135.2   194.0   234.7  273.8   561.0

> wilcox.test(n.c ~ (y.c == 0))

      Wilcoxon rank sum test with continuity correction

data:  n.c by y.c == 0
W = 156, p-value = 0.9113
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(n.t ~ (y.t == 0))
```

Wilcoxon rank sum test

```
data: n.t by y.t == 0
W = 103, p-value = 0.6045
alternative hypothesis: true location shift is not equal to 0
```

If we simulate as if they are Binomial, what would happen? On average we can see control arm is zero inflated, but for treatment arm, its zero deflated, and ratio of number of trials with zero event is far below the observed 20/6!

```
> z.c <- z.t <- r <- NULL
> for (i in 1:5000) {
+   y.0 <- rbinom(length(n.c), n.c, pi.c)
+   y.1 <- rbinom(length(n.t), n.t, pi.t)
+   z.c <- c(z.c, sum(y.0 == 0))
+   z.t <- c(z.t, sum(y.1 == 0))
+   r <- c(r, sum(y.0 == 0)/sum(y.1 == 0))
+ }
> summary(z.c)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  9.00  18.00   20.00   20.35  22.00   30.00

> summary(z.t)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  4.00  12.00   14.00   13.87  16.00   23.00

> summary(r)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.550  1.267   1.467   1.524  1.727   5.750
```

Next we simulate according to the following strategy: for control arm, generate $Z_i^c \sim \text{Bernoulli}(p_c)$, if $Z_i^c = 0$, then $Y_i^c = 0$, if $Z_i^c = 1$, then $Y_i^c \sim ZTB(\pi_c)$, where $ZTB(\cdot)$ denotes the zero-truncated Binomial distribution. The responses for treatment arm $\{Y_i^T, i = 1, \dots, n\}$ can be generated in a similar fashion. In the following, 40% trials have at least one event in the control arms, while 75% trials have at least one event in the treatment arm. For trials with at least one event in the control arm, the number of events in the control arm follows a zero-truncated Binomial distribution with probability 0.0083, for trials with at least one event in the treatment arm, the number of events in the treatment arm follows zero-truncated Binomial with probability 0.0125. The true risk ratio is 1.5. It is easily seen that data generated this way are more closely resemble the observed data.

The following simulation shows all methods give biased results. Peto method, Mantel-Haenszel method without zero correction (as the authors did in the NEJM paper) seriously over-estimate the treatment effect. The results are better with zero correction, where Peto method gives the best results. Note the results depend on the distributional assumption on the observed data, thus limiting its generalizability.

```
> n.c <- rosig$n.c[indx < 41]
> n.t <- rosig$n.t[indx < 41]
> y.c <- rosig$e.mi.c[indx < 41]
> y.t <- rosig$e.mi.t[indx < 41]
> p.c <- sum(y.c > 0)/length(y.c)
> p.t <- sum(y.t > 0)/length(y.t)
> pi.c <- sum(y.c[y.c > 0])/sum(n.c[y.c > 0])
> r <- 1.5
> pi.t <- pi.c * r
> N <- length(n.c)
> nsim <- 1000
> out <- NULL
```

```

> OR.Peto <- OR.MH <- OR.Peto.zc <- OR.MH.zc <- NULL
> for (i in 1:nsim) {
+   z.c <- rbinom(N, 1, p.c)
+   z.t <- rbinom(N, 1, p.t)
+   y.c <- y.t <- rep(0, N)
+   y.c[z.c == 1] <- rztb(sum(z.c), n.c[z.c == 1], pi.c)
+   y.t[z.t == 1] <- rztb(sum(z.t), n.t[z.t == 1], pi.t)
+   dat <- data.frame(study = c(1:N), n.t, y.t, n.c, y.c)
+   out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+     method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)
+   OR.Peto <- c(OR.Peto, exp(summary(out)$fixed$TE))
+   out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+     method = "MH", sm = "OR", incr = 0, allstudies = TRUE)
+   OR.MH <- c(OR.MH, exp(summary(out)$fixed$TE))
+   ind.1 <- y.t + y.c
+   ind.2 <- y.t * y.c
+   dat$n.t[ind.1 > 0 & ind.2 == 0] <- dat$n.t[ind.1 > 0 & ind.2 ==
+     0] + 1
+   dat$n.c[ind.1 > 0 & ind.2 == 0] <- dat$n.c[ind.1 > 0 & ind.2 ==
+     0] + 1
+   dat$y.t[ind.1 > 0 & ind.2 == 0] <- dat$y.t[ind.1 > 0 & ind.2 ==
+     0] + 0.5
+   dat$y.c[ind.1 > 0 & ind.2 == 0] <- dat$y.c[ind.1 > 0 & ind.2 ==
+     0] + 0.5
+   out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+     method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)
+   OR.Peto.zc <- c(OR.Peto.zc, exp(summary(out)$fixed$TE))
+   out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+     method = "MH", sm = "OR", incr = 0, allstudies = TRUE)
+   OR.MH.zc <- c(OR.MH.zc, exp(summary(out)$fixed$TE))
+ }
> summary(OR.Peto)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.131  1.825   2.059   2.113  2.375   3.989

> summary(OR.MH)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.135  1.924   2.289   2.450  2.810   8.931

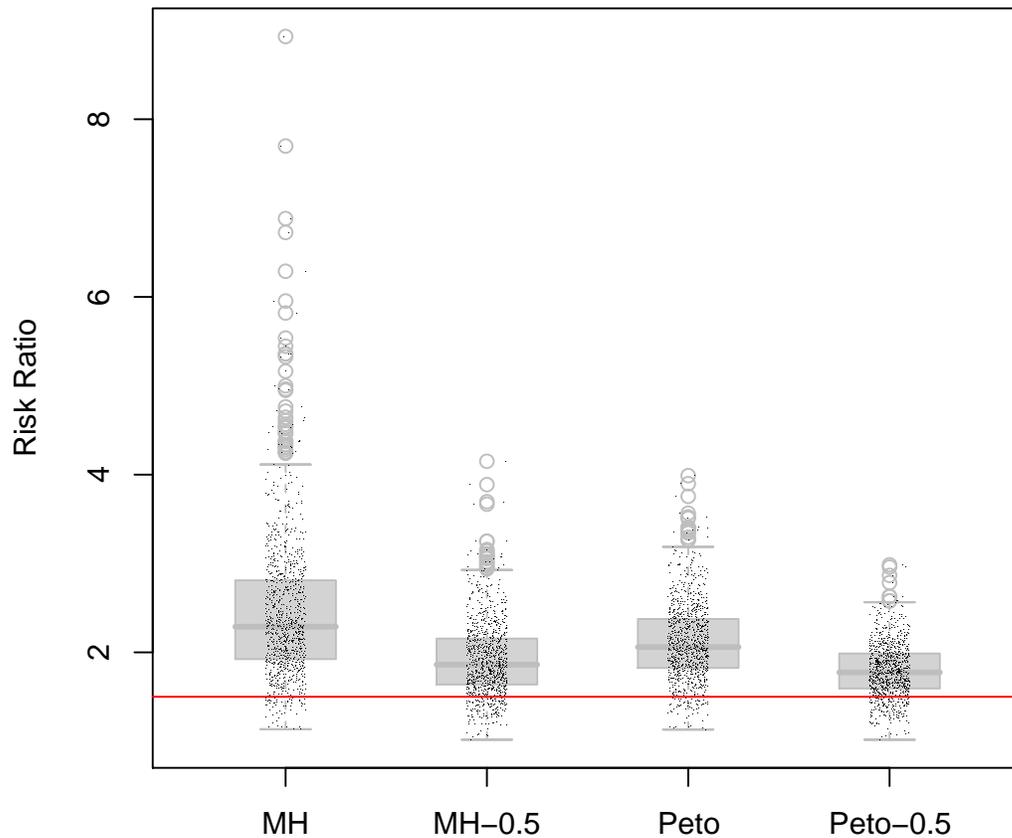
> summary(OR.Peto.zc)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.017  1.593   1.775   1.795  1.987   2.985

> summary(OR.MH.zc)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.017  1.639   1.861   1.920  2.155   4.151

```



I have to admit that the mixture of zero state and zero-truncated binomial is a rather awkward. A binomial model with sample size dependent baseline rates may be more appropriate (suggested by Thomas Lumley). It is under further investigation. One question is that how it is related to the random effects model.

4.1 An interesting problem

Now consider the following problem: suppose there are 20 randomized-controlled studies. For each study, the total sample size is 200, with 100 in the treatment arm and 100 in the control arm. Suppose 16 of the studies each had 1 event in the treatment arm and 0 event in the control arm, the left 4 studies each had 4 events in the control arm and 0 event in the treatment arm. The crude overall event rates are the same for the two arms: 16/4000. The question is: is there a significant difference in the event rate between the two arms? On one hand, you have a majority of studies had one event in the treatment arm, on the other other hand, you have more events in the control arm for a minority of the studies. Let's see how Peto method perform under this scenario. For fixed effects model, Peto method with and without zero correction gave similar results, but random effects model gave almost significant results.

```
> n.t <- n.c <- rep(100, 20)
> y.t <- c(rep(1, 16), rep(0, 4))
> y.c <- c(rep(0, 16), rep(4, 4))
> dat <- data.frame(study = c(1:N), n.t, y.t, n.c, y.c)
> out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
```

```
+ method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)
> summary(out)
```

Number of trials combined: 40

	OR	95%-CI	z	p.value
Fixed effects model	1.0000	[0.6115; 1.6353]	0.0000	1
Random effects model	1.8357	[0.9417; 3.5782]	1.7837	0.0745

Quantifying heterogeneity:

tau² = 1.6698; H = 1.29 [1.06; 1.56]; I² = 39.5% [11.4%; 58.7%]

Test of heterogeneity:

Q	d.f.	p.value
64.49	39	0.0063

Method: Peto method

```
> out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+ method = "MH", sm = "OR", incr = 0, allstudies = TRUE)
> summary(out)
```

Number of trials combined: 40

	OR	95%-CI	z	p.value
Fixed effects model	1	[0.6126; 1.6323]	0	1
Random effects model	NaN	[NaN; NaN]	NaN	--

Quantifying heterogeneity:

tau² = 0; H = 1 [1; 1]; I² = 0% [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
0	39	1

Method: Mantel-Haenszel method

```
> out <- metabin(y.t, n.t, y.c, n.c, studlab = study, data = dat,
+ method = "Peto", sm = "OR", incr = 0.5, allstudies = TRUE)
> summary(out)
```

Number of trials combined: 40

	OR	95%-CI	z	p.value
Fixed effects model	1.0000	[0.6115; 1.6353]	0.0000	1
Random effects model	1.8357	[0.9417; 3.5782]	1.7837	0.0745

Quantifying heterogeneity:

tau² = 1.6698; H = 1.29 [1.06; 1.56]; I² = 39.5% [11.4%; 58.7%]

Test of heterogeneity:

Q	d.f.	p.value
64.49	39	0.0063

Method: Peto method

5 Bayesian analysis

We consider only the 40 small studies. Among these 40 studies, 4 studies are dropped due to lack of events in both arms.

```
> indx <- c(1:42)
> dat <- rosig
> ind.1 <- dat$e.mi.t + dat$e.mi.c
> ind.2 <- dat$e.mi.t * dat$e.mi.c
> k <- 36
> r.c <- dat$e.mi.c[indx < 41 & ind.1 > 0]
> r.t <- dat$e.mi.t[indx < 41 & ind.1 > 0]
> n.c <- dat$n.c[indx < 41 & ind.1 > 0]
> n.t <- dat$n.t[indx < 41 & ind.1 > 0]
> datta <- c("k", "r.t", "n.t", "r.c", "n.c")
> delta <- rep(1.43, k)
> theta <- 1.43
> tau <- 1
> pars <- c("delta", "theta", "tau")
> dump.jags(datta, "meta.dat")
```

```
[1] TRUE
```

```
> dump.jags(pars, "meta.in")
```

```
[1] TRUE
```

```
> system("jags OR-a.cmd")
```

This Bayesian model assumes study specific event rates, which fits better to the observed raw data. The BUGS code program is as follows,

```
# Odds ratio scale: delta = log odds ratio
# fixed effect: flat prior on pi.c

var r.c[k], n.c[k], r.t[k], n.t[k], pi.c[k], pi.t[k],
    mu[k], delta[k], theta, tau, precision.tau;
model {
  for (i in 1:k) {
    r.c[i] ~ dbin(pi.c[i], n.c[i]);
    r.t[i] ~ dbin(pi.t[i], n.t[i]);
    mu[i] <- logit(pi.c[i]);
    logit(pi.t[i]) <- mu[i] + delta[i];
    delta[i] ~ dnorm(theta, precision.tau);
    pi.c[i] ~ dunif(0, 0.01); # pi.c are unrelated, fixed effect
  }
  theta ~ dnorm(0, 0.1);
  precision.tau <- 1/(tau*tau);
  tau ~ dunif(0, 2);
}
```

Now we look at the posterior summaries. The posterior mean of odds ratio is 1.058, posterior median is 1.04, 95% posterior confidence interval is (0.669, 1.542).

```
> require(coda)
```

```
[1] TRUE
```

```
> out <- read.jags("rosig.out")
```

```
Abstracting theta ... 10000 valid values
Abstracting tau ... 10000 valid values
```

```
> summary(out)
```

```
Iterations = 50001:99996
Thinning interval = 5
Number of chains = 1
Sample size per chain = 10000
```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
theta	0.03298	0.2129	0.002129	0.006265
tau	0.43555	0.2592	0.002592	0.017121

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
theta	-0.40801	-0.0991	0.03504	0.1745	0.4428
tau	0.03358	0.2311	0.41440	0.6093	0.9865

```
> summary(exp(out[, 1]))
```

```
Iterations = 50001:99996
Thinning interval = 5
Number of chains = 1
Sample size per chain = 10000
```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
	1.057012	0.225032	0.002250	0.006529

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
	0.6650	0.9056	1.0357	1.1906	1.5570

For all studies combined excluding 4 studies with zero event in both arms ($k = 38$), the posterior mean is 1.208, posterior median is 1.194, 95% confidence interval is (0.782, 1.721).

```
> k <- 38
> r.c <- dat$e.mi.c[ind.1 > 0]
> r.t <- dat$e.mi.t[ind.1 > 0]
> n.c <- dat$n.c[ind.1 > 0]
> n.t <- dat$n.t[ind.1 > 0]
> datta <- c("k", "r.t", "n.t", "r.c", "n.c")
> delta <- rep(1.43, k)
> theta <- 1.43
> tau <- 1
> pars <- c("delta", "theta", "tau")
> dump.jags(datta, "meta.dat")
```

```
[1] TRUE
```

```

> dump.jags(pars, "meta.in")

[1] TRUE

> system("jags OR-a.cmd")
> out <- read.jags("rosig.out")

Abstracting theta ... 10000 valid values
Abstracting tau ... 10000 valid values

> summary(out)

Iterations = 50001:99996
Thinning interval = 5
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

      Mean      SD Naive SE Time-series SE
theta 0.1781 0.1941 0.001941      0.006246
tau   0.4145 0.2276 0.002276      0.011807

2. Quantiles for each variable:

      2.5%    25%    50%    75%   97.5%
theta -0.24014 0.05546 0.1910 0.3163 0.5233
tau    0.04957 0.24207 0.4002 0.5598 0.8992

> summary(exp(out[, 1]))

Iterations = 50001:99996
Thinning interval = 5
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

      Mean      SD      Naive SE Time-series SE
1.217173    0.230406    0.002304    0.007612

2. Quantiles for each variable:

 2.5%  25%  50%  75%  97.5%
0.7865 1.0570 1.2105 1.3721 1.6876

```

Note even though 4 studies have no event in both arms, the more than 3000 patients in these studies still contains important information about the effect size and precision of the estimates. So next we consider all $n = 42$ studies. The posterior mean of overall OR is 1.145 with 95% confidence interval (0.71, 1.64). There is no statistical significance when all the studies are combined.

```

> library(R2WinBUGS)
> metamodel <- function() {
+   for (i in 1:k) {
+     r.c[i] ~ dbin(pi.c[i], n.c[i])
+     r.t[i] ~ dbin(pi.t[i], n.t[i])

```

```

+     mu[i] <- logit(pi.c[i])
+     logit(pi.t[i]) <- mu[i] + delta[i]
+     delta[i] ~ dnorm(theta, precision.tau)
+     pi.c[i] ~ dunif(0, 0.01)
+   }
+   theta ~ dnorm(0, 0.1)
+   precision.tau <- 1/(tau * tau)
+   tau ~ dunif(0, 2)
+ }
> r.c <- dat$e.mi.c
> r.t <- dat$e.mi.t
> n.c <- dat$n.c
> n.t <- dat$n.t
> k <- length(r.c)
> datta <- list("k", "r.t", "n.t", "r.c", "n.c")
> delta <- rep(1.43, k)
> theta <- 1.43
> tau <- 1
> inits <- list(delta, theta, tau)
> pars <- c("delta", "theta")
> write.model(metamodel, "metafixed.bug")
> meta.sim <- bugs(datta, inits, pars, "metafixed.bug", n.chains = 3,
+   n.iter = 5000, bugs.directory = "/home/zhouc2/.wine/drive_c/Program Files/WinBUGS14/",
+   working.directory = NULL, clearWD = TRUE, useWINE = TRUE,
+   WINE = "wine", newWINE = TRUE, WINEPATH = "/usr/bin/winepath")

> print(meta.sim)

```

Inference for Bugs model at "metafixed.bug", fit using winbugs,
 3 chains, each with 5000 iterations (first 2500 discarded), n.thin = 7
 n.sims = 1074 iterations saved

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
delta[1]	0.1	0.4	-0.9	-0.1	0.2	0.4	0.9	1	450
delta[2]	0.1	0.4	-1.0	-0.2	0.2	0.4	0.8	1	320
delta[3]	-0.1	0.5	-1.3	-0.4	0.0	0.3	0.7	1	99
delta[4]	-0.1	0.6	-1.4	-0.4	0.0	0.3	0.7	1	150
delta[5]	0.1	0.5	-1.0	-0.2	0.2	0.4	0.9	1	190
delta[6]	0.0	0.6	-1.3	-0.2	0.2	0.4	1.0	1	170
delta[7]	0.2	0.5	-0.9	0.0	0.2	0.5	1.2	1	1100
delta[8]	0.7	0.5	0.0	0.3	0.6	1.0	1.8	1	190
delta[9]	0.1	0.5	-1.1	-0.2	0.2	0.4	1.0	1	170
delta[10]	0.1	0.5	-1.1	-0.2	0.2	0.4	1.0	1	290
delta[11]	-0.1	0.6	-1.5	-0.4	0.0	0.3	0.7	1	120
delta[12]	0.1	0.5	-1.0	-0.2	0.2	0.4	1.0	1	600
delta[13]	0.2	0.5	-0.8	-0.1	0.2	0.4	1.0	1	200
delta[14]	0.2	0.5	-0.9	-0.1	0.2	0.4	1.0	1	570
delta[15]	0.1	0.5	-1.0	-0.2	0.1	0.4	0.8	1	210
delta[16]	0.0	0.5	-1.1	-0.3	0.1	0.3	0.9	1	250
delta[17]	0.1	0.5	-0.9	-0.1	0.2	0.4	0.9	1	240
delta[18]	0.2	0.5	-0.7	0.0	0.3	0.5	1.2	1	600
delta[19]	0.4	0.4	-0.4	0.2	0.4	0.7	1.4	1	1100
delta[20]	0.0	0.6	-1.3	-0.3	0.1	0.3	1.0	1	160
delta[21]	0.0	0.6	-1.4	-0.3	0.1	0.3	0.8	1	240
delta[22]	0.0	0.5	-1.3	-0.3	0.1	0.3	0.8	1	120
delta[23]	0.2	0.5	-0.8	0.0	0.3	0.5	1.3	1	550
delta[24]	0.3	0.5	-0.8	0.0	0.3	0.5	1.4	1	1100

delta[25]	-0.2	0.6	-1.6	-0.6	-0.2	0.2	0.6	1	78
delta[26]	0.3	0.4	-0.7	0.1	0.3	0.5	1.2	1	1100
delta[27]	0.2	0.5	-0.9	-0.1	0.2	0.5	1.1	1	710
delta[28]	0.1	0.5	-1.0	-0.1	0.2	0.4	1.1	1	100
delta[29]	0.2	0.5	-0.7	0.0	0.2	0.5	1.1	1	410
delta[30]	0.2	0.5	-0.9	-0.1	0.2	0.4	1.2	1	490
delta[31]	0.0	0.5	-1.2	-0.2	0.1	0.3	0.9	1	220
delta[32]	0.0	0.5	-1.2	-0.3	0.1	0.3	0.9	1	110
delta[33]	-0.1	0.6	-1.4	-0.4	0.1	0.3	0.8	1	61
delta[34]	0.0	0.5	-1.1	-0.2	0.1	0.4	0.8	1	240
delta[35]	0.1	0.5	-1.0	-0.2	0.2	0.4	1.1	1	92
delta[36]	0.1	0.5	-1.1	-0.1	0.2	0.4	1.0	1	550
delta[37]	0.1	0.5	-1.1	-0.2	0.1	0.4	1.0	1	510
delta[38]	0.0	0.6	-1.3	-0.3	0.1	0.4	1.0	1	240
delta[39]	0.0	0.5	-1.1	-0.3	0.1	0.3	0.9	1	230
delta[40]	0.0	0.5	-1.2	-0.3	0.1	0.3	1.0	1	220
delta[41]	0.3	0.3	-0.3	0.1	0.3	0.5	0.9	1	520
delta[42]	0.5	0.2	0.2	0.4	0.5	0.7	0.9	1	790
theta	0.1	0.2	-0.3	0.0	0.1	0.3	0.5	1	56
deviance	210.5	9.0	194.3	204.1	210.2	216.5	228.2	1	1100

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

pD = 23.0 and DIC = 233.5 (using the rule, pD = Dbar-Dhat)
DIC is an estimate of expected predictive error (lower deviance is better).

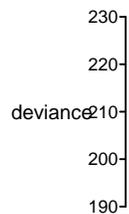
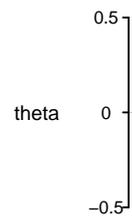
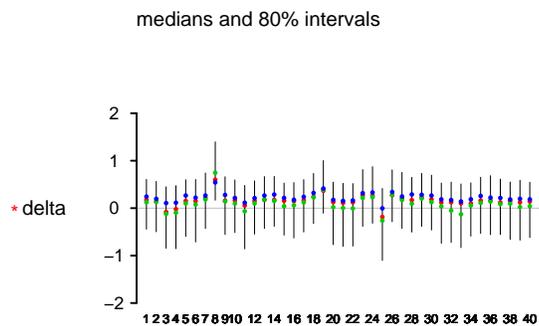
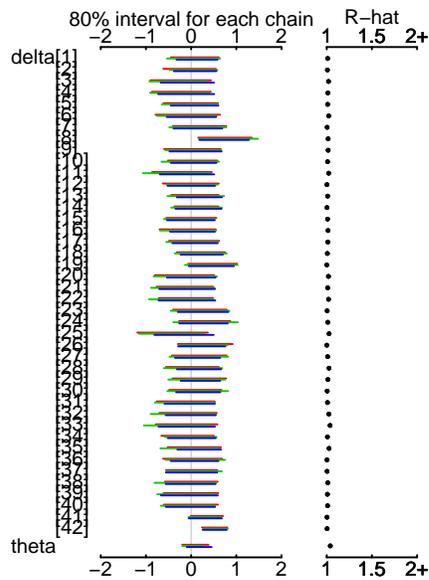
```
> summary(exp(meta.sim$sims.list$theta))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4483	0.9702	1.1430	1.1450	1.2970	1.8730

```
> quantile(exp(meta.sim$sims.list$theta), probs = c(0.025, 0.975))
```

2.5%	97.5%
0.7066305	1.6434621

Bugs model at "metafixed.bug", fit using winbugs, 3 chains, each with 5000 iterations (first 2500 discarded)



6 Sensitivity analysis with zero-correction

We now re-run the analysis with zero correction to avoid the imbalance due to the sparseness in the data. We will run both fixed effects model and random effects model.

We first exclude 4 trials with no event in both arms. When no event occurs in either arm of a trial, the trial odds ratio is undefined, as the trial provides no information about either the likely direction or magnitude of the effect. For trials with zero event in only one arm, we apply the strategy used by [4]. We add 0.5 to all cells for all trials that have zero events in either arm.

6.1 MI event

```
> indx <- c(1:42)
> dat <- rosig
> ind.1 <- dat$e.mi.t + dat$e.mi.c
> ind.2 <- dat$e.mi.t * dat$e.mi.c
> y <- dat$n.t[ind.1 > 0 & indx < 41]
> sum(y)
```

[1] 10285

```

> y <- dat$n.c[ind.1 > 0 & indx < 41]
> sum(y)

[1] 6106

> dat <- rosig
> dat$n.t[ind.1 > 0 & ind.2 == 0] <- dat$n.t[ind.1 > 0 & ind.2 ==
+ 0] + 1
> dat$n.c[ind.1 > 0 & ind.2 == 0] <- dat$n.c[ind.1 > 0 & ind.2 ==
+ 0] + 1
> dat$e.mi.t[ind.1 > 0 & ind.2 == 0] <- dat$e.mi.t[ind.1 > 0 &
+ ind.2 == 0] + 0.5
> dat$e.mi.c[ind.1 > 0 & ind.2 == 0] <- dat$e.mi.c[ind.1 > 0 &
+ ind.2 == 0] + 0.5

```

Now we look at the small studies combined. The pooled estimate of OR based on 36 small trials is 1.17, 95% CI is (0.76, 1.78), P value is 0.48. The effect size is much smaller with zero correction than those reported in ??.

```

> meta.1 <- metabin(e.mi.t, n.t, e.mi.c, n.c, studlab = study,
+ subset = (ind.1[indx < 41] > 0), data = dat[indx < 41, ],
+ method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)
> summary(meta.1)

```

Number of trials combined: 36

	OR	95%-CI	z	p.value
Fixed effects model	1.1659	[0.7616; 1.7849]	0.7065	0.4799
Random effects model	1.1659	[0.7616; 1.7849]	0.7065	0.4799

Quantifying heterogeneity:

tau² = 0; I² = 0% [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
18.2	35	0.9914

Method: Peto method

We next calculate the pooled estimate of OR using all the trials. We confirm the results for DREAM and ADOPT trials. The overall OR is estimated to be 1.28, with 95% CI (0.95, 1.74) and P value 0.11. Based on these results, the meta-analysis shows there is NO statistically significant association between Rosiglitazone and increased risk of myocardial infarction, quite contrary to what these authors claimed!

```

> meta.2 <- metabin(e.mi.t, n.t, e.mi.c, n.c, studlab = study,
+ subset = (ind.1 > 0), data = dat, method = "Peto", sm = "OR",
+ incr = 0, allstudies = TRUE)
> meta.2

```

	OR	95%-CI	%W(fixed)	%W(random)
49653/011	2.1051	[0.1892; 23.4266]	1.57	1.57
49653/020	1.0585	[0.0977; 11.4661]	1.61	1.61
49653/024	0.1388	[0.0041; 4.6633]	0.74	0.74
49653/093	0.1594	[0.0085; 2.9890]	1.06	1.06
49653/094	1.4620	[0.0771; 27.7131]	1.06	1.06
100684	0.3966	[0.0244; 6.4551]	1.17	1.17
49653/143	2.7972	[0.1739; 44.9831]	1.18	1.18
49653/211	2.5039	[0.5575; 11.2460]	4.05	4.05

49653/284	2.7361	[0.1708; 43.8228]	1.19	1.19
712753/008	1.3977	[0.0719; 27.1657]	1.04	1.04
AVM100264	0.3772	[0.0235; 6.0456]	1.19	1.19
BRL49653C/185	1.2476	[0.0743; 20.9538]	1.15	1.15
BRL49653C/334	1.9595	[0.2030; 18.9146]	1.78	1.78
BRL49653/347	2.1487	[0.1954; 23.6314]	1.59	1.59
49653/015	1.0025	[0.0906; 11.0900]	1.58	1.58
49653/079	0.4972	[0.0267; 9.2572]	1.07	1.07
49653/080	0.4850	[0.0499; 4.7191]	1.76	1.76
49653/082	2.1496	[0.1945; 23.7533]	1.58	1.58
49653/085	2.7685	[0.3858; 19.8662]	2.35	2.35
49653/097	0.3604	[0.0224; 5.7948]	1.18	1.18
49653/125	0.3626	[0.0226; 5.8210]	1.19	1.19
49653/127	2.8402	[0.1754; 45.9806]	1.18	1.18
49653/128	2.6845	[0.1649; 43.7054]	1.17	1.17
49653/134	0.1022	[0.0092; 1.1370]	1.57	1.57
49653/135	0.6362	[0.1084; 3.7319]	2.92	2.92
49653/136	2.6366	[0.1641; 42.3701]	1.18	1.18
49653/145	2.8551	[0.1779; 45.8131]	1.19	1.19
49653/147	2.7032	[0.1678; 43.5589]	1.18	1.18
49653/162	2.7914	[0.1739; 44.8181]	1.18	1.18
49653/330	0.9664	[0.0383; 24.4111]	0.88	0.88
49653/137	0.4629	[0.0477; 4.4888]	1.77	1.77
SB-712753/002	2.6481	[0.1652; 42.4511]	1.19	1.19
SB-712753/003	2.9196	[0.1818; 46.8763]	1.18	1.18
SB-712753/007	1.4358	[0.0751; 27.4614]	1.05	1.05
49653/132	0.7488	[0.0238; 23.5196]	0.77	0.77
AVA100193	0.9486	[0.0369; 24.3769]	0.87	0.87
DREAM	1.6517	[0.7407; 3.6830]	14.20	14.20
ADOPT	1.3294	[0.8002; 2.2086]	35.44	35.44

Number of trials combined: 38

	OR	95%-CI	z	p.value
Fixed effects model	1.2834	[0.9486; 1.7361]	1.618	0.1057
Random effects model	1.2834	[0.9486; 1.7361]	1.618	0.1057

Quantifying heterogeneity:

$\tau^2 = 0$; $I^2 = 0\%$ [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
18.79	37	0.9944

Method: Peto method

We now run random effects model and fixed effects model based on MH method. The results are very similar to those from Peto method. They still show there is no strong evidence to suggest an association between Rosiglitazone and increased MI risk.

```
> meta.3 <- metabin(e.mi.t, n.t, e.mi.c, n.c, studlab = study,
+ subset = (ind.1 > 0), data = dat, method = "MH", sm = "OR",
+ incr = 0, allstudies = TRUE)
> summary(meta.3)
```

Number of trials combined: 38

	OR	95%-CI	z	p.value
Fixed effects model	1.2776	[0.9463; 1.7248]	1.5994	0.1097
Random effects model	1.2856	[0.9398; 1.7587]	1.5713	0.1161

Quantifying heterogeneity:
 $\tau^2 = 0$; $I^2 = 0\%$ [0%; 0%]

Test of heterogeneity:
 Q d.f. p.value
 16.22 37 0.9988

Method: Mantel-Haenszel method

As a double check, we re-run the fixed effects model and random effects model using the function *meta.MH* and *meta.DSL* in the *rmeta* package.

```
> a <- meta.MH(n.t, n.c, e.mi.t, e.mi.c, data = dat, names = study,
+ subset = (ind.1 > 0))
> summary(a)
```

```
Fixed effects ( Mantel-Haenszel ) meta-analysis
Call: meta.MH(ntrt = n.t, nctrl = n.c, ptrt = e.mi.t, pctrl = e.mi.c,
names = study, data = dat, subset = (ind.1 > 0))
```

```
-----
OR (lower 95% upper)
49653/011 2.48 0.12 51.99
49653/020 1.06 0.10 11.75
49653/024 0.24 0.01 3.82
49653/093 0.17 0.01 4.19
49653/094 1.51 0.06 37.35
100684 0.36 0.01 8.98
49653/143 3.10 0.13 76.84
49653/211 2.67 0.51 14.04
49653/284 3.02 0.12 74.45
712753/008 1.43 0.06 35.43
AVM100264 0.34 0.01 8.41
BRL49653C/185 1.27 0.06 26.58
BRL49653C/334 2.01 0.18 22.34
BRL49653/347 2.55 0.12 53.38
49653/015 1.00 0.09 11.12
49653/079 0.52 0.03 8.39
49653/080 0.47 0.04 5.28
49653/082 2.55 0.12 53.66
49653/085 3.07 0.32 29.85
49653/097 0.33 0.01 8.06
49653/125 0.33 0.01 8.10
49653/127 3.16 0.13 79.27
49653/128 3.00 0.12 75.96
49653/134 0.10 0.00 2.04
49653/135 0.63 0.10 3.85
49653/136 2.92 0.12 72.24
49653/145 3.16 0.13 77.87
49653/147 3.00 0.12 74.65
49653/162 3.09 0.12 76.38
49653/330 0.97 0.04 23.78
49653/137 0.45 0.04 5.01
SB-712753/002 2.93 0.12 72.15
```

SB-712753/003	3.22	0.13	79.52
SB-712753/007	1.48	0.06	36.50
49653/132	0.76	0.03	18.89
AVA100193	0.95	0.04	23.45
DREAM	1.67	0.73	3.82
ADOPT	1.32	0.81	2.15

Mantel-Haenszel OR =1.28 95% CI (0.95,1.72)
Test for heterogeneity: $X^2(37) = 16.22$ (p-value 0.9988)

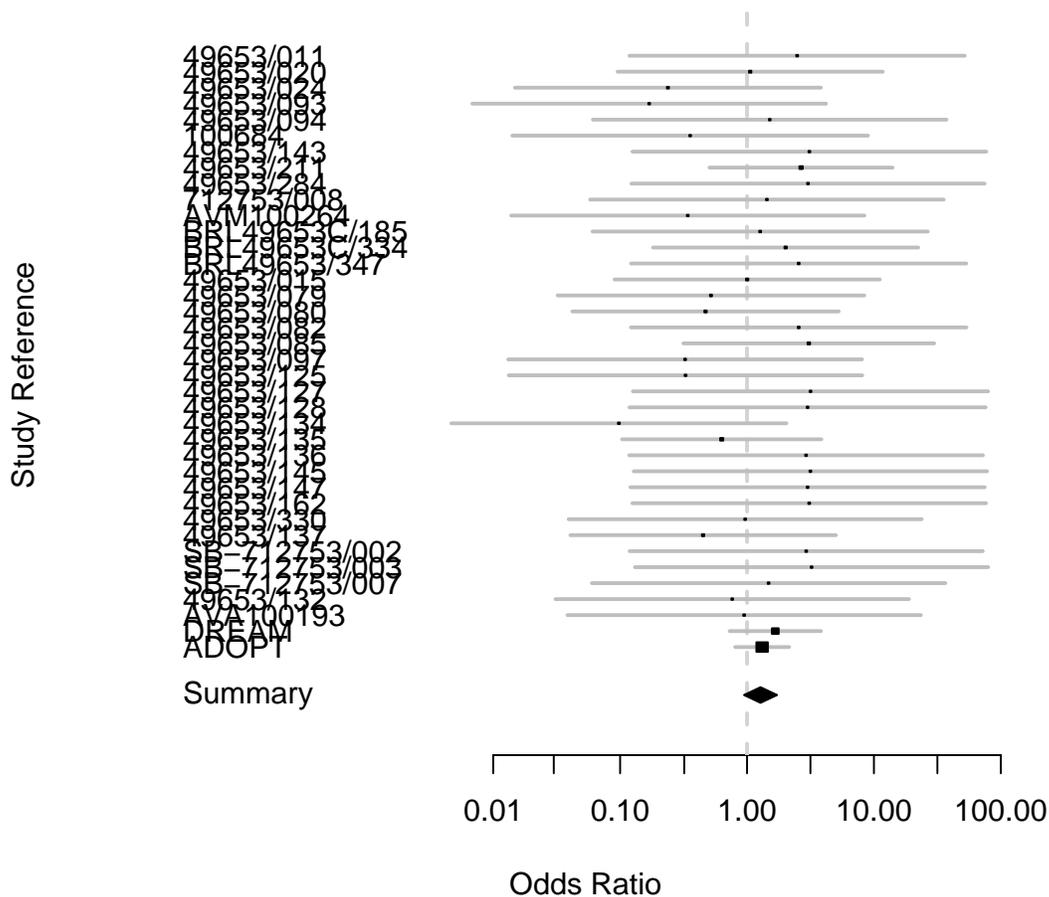
```
> b <- meta.DSL(n.t, n.c, e.mi.t, e.mi.c, data = dat, names = study,
+ subset = (ind.1 > 0))
> summary(b)
```

Random effects (DerSimonian-Laird) meta-analysis
Call: meta.DSL(n.trt = n.t, n.ctrl = n.c, ptrt = e.mi.t, pctrl = e.mi.c,
names = study, data = dat, subset = (ind.1 > 0))

	OR	(lower	95% upper)
49653/011	2.48	0.12	51.99
49653/020	1.06	0.10	11.75
49653/024	0.24	0.01	3.82
49653/093	0.17	0.01	4.19
49653/094	1.51	0.06	37.35
100684	0.36	0.01	8.98
49653/143	3.10	0.13	76.84
49653/211	2.67	0.51	14.04
49653/284	3.02	0.12	74.45
712753/008	1.43	0.06	35.43
AVM100264	0.34	0.01	8.41
BRL49653C/185	1.27	0.06	26.58
BRL49653C/334	2.01	0.18	22.34
BRL49653/347	2.55	0.12	53.38
49653/015	1.00	0.09	11.12
49653/079	0.52	0.03	8.39
49653/080	0.47	0.04	5.28
49653/082	2.55	0.12	53.66
49653/085	3.07	0.32	29.85
49653/097	0.33	0.01	8.06
49653/125	0.33	0.01	8.10
49653/127	3.16	0.13	79.27
49653/128	3.00	0.12	75.96
49653/134	0.10	0.00	2.04
49653/135	0.63	0.10	3.85
49653/136	2.92	0.12	72.24
49653/145	3.16	0.13	77.87
49653/147	3.00	0.12	74.65
49653/162	3.09	0.12	76.38
49653/330	0.97	0.04	23.78
49653/137	0.45	0.04	5.01
SB-712753/002	2.93	0.12	72.15
SB-712753/003	3.22	0.13	79.52
SB-712753/007	1.48	0.06	36.50
49653/132	0.76	0.03	18.89
AVA100193	0.95	0.04	23.45
DREAM	1.67	0.73	3.82

ADOPT 1.32 0.81 2.15

 SummaryOR= 1.29 95% CI (0.94,1.76)
 Test for heterogeneity: X²(37) = 16.22 (p-value 0.9988)
 Estimated random effects variance: 0



6.2 Death with cardiovascular causes

We now repeat the analysis for event of death with cardiovascular causes. First again let's try to duplicate what the authors did.

The following results for the small trials combined are in agreement with those reported in the paper.

```
> meta.c <- metabin(e.death.t, n.t, e.death.c, n.c, studlab = study,
+ data = rosig[indx < 41, ], method = "Peto", sm = "OR", incr = 0,
+ allstudies = TRUE)
> summary(meta.c)
```

Number of trials combined: 40

	OR	95%-CI	z	p.value
Fixed effects model	2.3949	[1.1685; 4.9085]	2.3853	0.0171

```
Random effects model 2.3949 [1.1685; 4.9085] 2.3853 0.0171
```

```
Quantifying heterogeneity:  
tau^2 = 0; H = 1 [1; 1]; I^2 = 0% [0%; 0%]
```

```
Test of heterogeneity:  
  Q d.f. p.value  
8.35  39      1
```

```
Method: Peto method
```

For all studies combined, we have duplicated the results as well. The overall OR is 1.64 with 95% CI (0.98, 2.74) and P value 0.06.

```
> meta.d <- metabin(e.death.t, n.t, e.death.c, n.c, studlab = study,  
+ data = rosig[, ], method = "Peto", sm = "OR", incr = 0, allstudies = TRUE)  
> summary(meta.d)
```

```
Number of trials combined: 42
```

	OR	95%-CI	z	p.value
Fixed effects model	1.64	[0.9801; 2.7443]	1.8834	0.0596
Random effects model	1.64	[0.9801; 2.7443]	1.8834	0.0596

```
Quantifying heterogeneity:  
tau^2 = 0; H = 1 [1; 1]; I^2 = 0% [0%; 0%]
```

```
Test of heterogeneity:  
  Q d.f. p.value  
10.75  41      1
```

```
Method: Peto method
```

Now we do zero correction. There are 19 trials with no event in either arms, thus are excluded from subsequent analyses. Among 17 trials with zero event in one of the arms, only 2 trials had zero event in the treatment arm, and the other 15 trials all had zero event in the control arm, even though the risk differences are small. Following similar argument, Peto method without zero correction will tend to overestimate the effect size with increased type I error.

```
> dat <- rosig  
> ind.1 <- dat$e.death.t + dat$e.death.c  
> ind.2 <- dat$e.death.t * dat$e.death.c  
> sum(dat$n.t[ind.1 > 0 & indx < 41])  
  
[1] 6845  
  
> sum(dat$n.c[ind.1 > 0 & indx < 41])  
  
[1] 3980  
  
> dat$n.t[ind.1 > 0 & ind.2 == 0] <- dat$n.t[ind.1 > 0 & ind.2 ==  
+ 0] + 1  
> dat$n.c[ind.1 > 0 & ind.2 == 0] <- dat$n.c[ind.1 > 0 & ind.2 ==  
+ 0] + 1  
> dat$e.death.t[ind.1 > 0 & ind.2 == 0] <- dat$e.death.t[ind.1 >  
+ 0 & ind.2 == 0] + 0.5  
> dat$e.death.c[ind.1 > 0 & ind.2 == 0] <- dat$e.death.c[ind.1 >  
+ 0 & ind.2 == 0] + 0.5
```

Second we combine small studies. Note there are 19 studies excluded due to lack of event in both arms. The overall OR is estimated to be 1.49 with 95% CI (0.83, 2.67) and P value 0.18.

```
> meta.4 <- metabin(e.death.t, n.t, e.death.c, n.c, studlab = study,
+ subset = (ind.1[indx < 41] > 0), data = dat[indx < 41, ],
+ method = "Peto", sm = "OR", allstudies = TRUE)
> summary(meta.4)
```

Number of trials combined: 21

	OR	95%-CI	z	p.value
Fixed effects model	1.4911	[0.8317; 2.6735]	1.3412	0.1799
Random effects model	1.4911	[0.8317; 2.6735]	1.3412	0.1799

Quantifying heterogeneity:

tau² = 0; H = 1 [1; 1]; I² = 0% [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
4.51	20	0.9999

Method: Peto method

Then we combine all the trials. The overall odds ratio is estimated to be 1.33 with 95% CI (0.83, 2.10) and P value 0.23. These results again differ dramatically from those reported in [2]. The results from random effects model are also almost identical to those from the fixed effects model.

```
> meta.5 <- metabin(e.death.t, n.t, e.death.c, n.c, studlab = study,
+ subset = (ind.1 > 0), data = dat, method = "Peto", sm = "OR",
+ incr = 0, allstudies = TRUE)
> meta.5
```

	OR	95%-CI	%W(fixed)	%W(random)
49653/011	1.4417	[0.0756; 27.4978]	2.41	2.41
49653/094	1.4620	[0.0771; 27.7131]	2.42	2.42
49653/211	1.5586	[0.2657; 9.1438]	6.70	6.70
AVM100264	2.0057	[0.2078; 19.3592]	4.08	4.08
BRL49653C/334	0.3685	[0.0230; 5.9067]	2.73	2.73
49653/015	2.1304	[0.1928; 23.5428]	3.64	3.64
49653/079	0.4972	[0.0267; 9.2572]	2.45	2.45
49653/082	1.4747	[0.0781; 27.8323]	2.43	2.43
49653/085	2.7478	[0.1710; 44.1522]	2.72	2.72
49653/095	1.4373	[0.0750; 27.5331]	2.41	2.41
49653/134	1.4375	[0.0753; 27.4312]	2.41	2.41
49653/135	1.8748	[0.1929; 18.2157]	4.06	4.06
49653/136	3.7008	[0.3818; 35.8677]	4.07	4.07
49653/145	2.8551	[0.1779; 45.8131]	2.72	2.72
49653/162	2.7914	[0.1739; 44.8181]	2.72	2.72
49653/330	0.9664	[0.0383; 24.4111]	2.01	2.01
49653/331	1.3552	[0.0686; 26.7742]	2.36	2.36
49653/137	0.3320	[0.0206; 5.3443]	2.72	2.72
SB-712753/002	2.6481	[0.1652; 42.4511]	2.73	2.73
49653/132	0.7488	[0.0238; 23.5196]	1.77	1.77
AVA100193	0.9486	[0.0369; 24.3769]	1.99	1.99
DREAM	1.1998	[0.5193; 2.7720]	29.93	29.93
ADOPT	0.8025	[0.1668; 3.8613]	8.50	8.50

Number of trials combined: 23

	OR	95%-CI	z	p.value
Fixed effects model	1.3255	[0.8383; 2.0958]	1.2056	0.228
Random effects model	1.3255	[0.8383; 2.0958]	1.2056	0.228

Quantifying heterogeneity:

$\tau^2 = 0$; $I^2 = 0\%$ [0%; 0%]

Test of heterogeneity:

Q	d.f.	p.value
5.11	22	0.9999

Method: Peto method

Based on above analysis, we again didn't find statistically significant association between Rosiglitazone and increased risk of death from cardiovascular causes.

7 Closing remarks

There have been heated discussion over the conclusion of this study in the medical community, but little attention has been given to the validity of their meta-analysis. Sensitivity analysis with zero-correction arrived at different conclusion. This cast some doubt onto the conclusions made in the paper. The Peto method these authors used is valid under certain assumptions, so we can't say the method or the results are wrong. But still with the discrepancies between the results from various methods, the findings are, to a large extent, not as conclusive as the paper suggested. Unfortunately the significance of the original paper and the buzz it has generated all depends on these calculated statistics, the analysis should have undergone more rigerous scrutiny before the public became aware of these results. It is a great example about how statistics plays a crucial role in today's medical research, and how statistics, if not used carefully, can mislead the public. In fact, I believe both the authors and the accompanying editorial tried to explain this viewpoint, but for some reason this message was neglected, at least by the media and the general public.

References

- [1] R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- [2] Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457-2471.
- [3] Psaty BM, Furberg CD. Rosiglitazone and cardiovascular risk (Editorial). *N Engl J Med*. 2007 Jun 14;356(24):2522-4.
- [4] Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26:53-77.
- [5] Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351-1375. [Erratum, *Stat Med* 2006;25:2700.]
- [6] Sutton A, Cooper N, Lambert P, Jones D, Abrams K, Sweeting M. Meta-analysis of rare and adverse event data. *Pharmacoeconomics Outcomes Res* 2002;2:367.