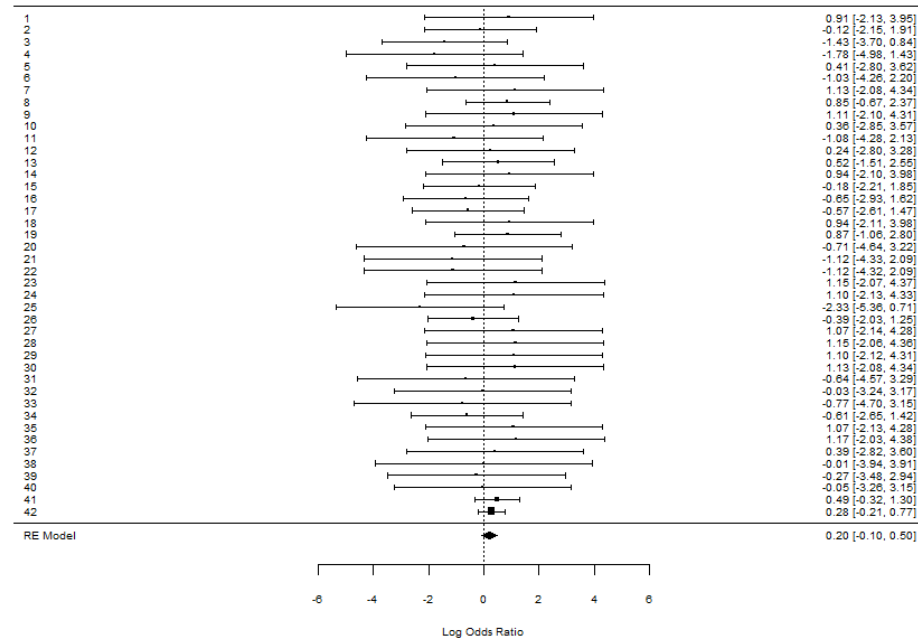Matt Isaac
A01515095
Biostatistics – Homework 7
Due: 4/25/18

1. The purpose of the NEJM article was to combine results from relevant studies about the effect of Rosiglitazone in a systematic way to investigate the effect of Rosiglitazone on the risk of death from cardiovascular causes and myocardial infarction.

2. As we combine effects from multiple studies, we can more clearly see if there is in fact a significant effect from the treatment in question. It increases our statistical power as well as allows us to determine if there is evidence of a publication bias.

3. A publication bias occurs when journals only publish studies that meet certain criteria. One might say that it occurs when journals only publish the "interesting" studies. Generally, when numerous journals only publish reports of large studies (i.e. large sample size) or studies that found significant effects (small p-values). Thus, the small studies that did not find significant do not get published. This makes it look like all the studies are finding significant effects/results when this might not be the case.

4. Conduct a random effects meta-analysis using the "+0.5" method we used in class, focusing on the odds ratio for myocardial infarction (heart attack) in the rosiglitazone group compared to the control group. Report the following:

   a. $\widehat{\theta}_i = \log\left(\frac{(b_i+0.5)(c_i+0.5)}{(a_i+0.5)(d_i+0.5)}\right)$ This is the percent increase (or decrease) in odds of event Y = 1 for Treatment group compared with control group.

b.

|	
1	0.91 [-2.13, 3.95]
2	-0.12 [-2.15, 1.91]
3	-1.43 [-3.70, 0.84]
4	-1.78 [-4.98, 1.43]
5	0.41 [-2.80, 3.62]
6	-1.03 [-4.26, 2.20]
7	1.13 [-2.08, 4.34]
8	0.85 [-0.67, 2.37]
9	1.11 [-2.10, 4.31]
10	0.36 [-2.85, 3.57]
11	-1.08 [-4.28, 2.13]
12	0.24 [-2.80, 3.28]
13	0.52 [-1.51, 2.55]
14	0.94 [-2.10, 3.98]
15	-0.18 [-2.21, 1.85]
16	-0.65 [-2.93, 1.62]
17	-0.57 [-2.61, 1.47]
18	0.94 [-2.11, 3.98]
19	0.87 [-1.06, 2.80]
20	-0.71 [-4.64, 3.22]
21	-1.12 [-4.33, 2.09]
22	-1.12 [-4.32, 2.09]
23	1.15 [-2.07, 4.37]
24	1.10 [-2.13, 4.33]
25	-2.33 [-5.36, 0.71]
26	-0.39 [-2.03, 1.25]
27	1.07 [-2.14, 4.28]
28	1.15 [-2.06, 4.36]
29	1.10 [-2.12, 4.31]
30	1.13 [-2.08, 4.34]
31	-0.64 [-4.57, 3.29]
32	-0.03 [-3.24, 3.17]
33	-0.77 [-4.70, 3.15]
34	-0.61 [-2.65, 1.42]
35	1.07 [-2.13, 4.28]
36	1.17 [-2.03, 4.38]
37	0.39 [-2.82, 3.60]
38	-0.01 [-3.94, 3.91]
39	-0.27 [-3.48, 2.94]
40	-0.05 [-3.26, 3.15]
41	0.49 [-0.32, 1.30]
42	0.28 [-0.21, 0.77]
RE Model	0.20 [-0.10, 0.50]

Log Odds Ratio

The most striking feature of the forest plot (above) is that every single confidence interval covers 0. In other words, in every study that was performed an odds ratio of 1 is a plausible outcome. This indicates that we have no evidence that the odds of an individual in the treatment group experiencing an myocardial infarction are any lower or higher than an individual in the control group experiencing a myocardial infarction. This is reflected in the diamond at the bottom of the plot, displaying the confidence band for the overall log odds ratio. Since this interval captures 0, it does not appear that we have significance in this situation.

c. $\widehat{OR} = e^{0.2009} = 1.22$

The odds ratio of 1.22 indicates that there is an estimated 22% increase in the odds of not having a myocardial infarction for patients who are given Rosiglitazone compared with patients who were given the control treatment. However, a significance test indicates that this difference is not significant.

d. Check for publication bias:
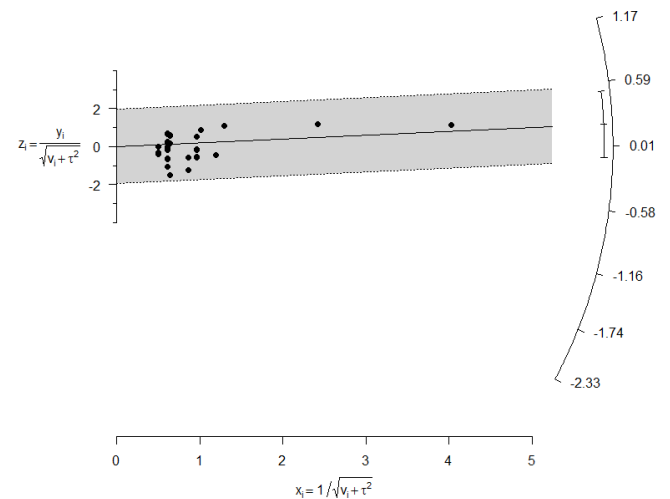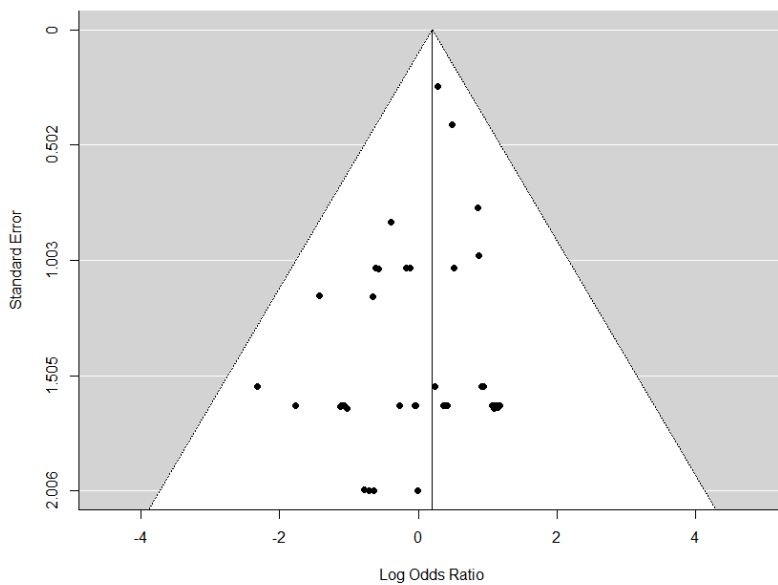
Numeric:

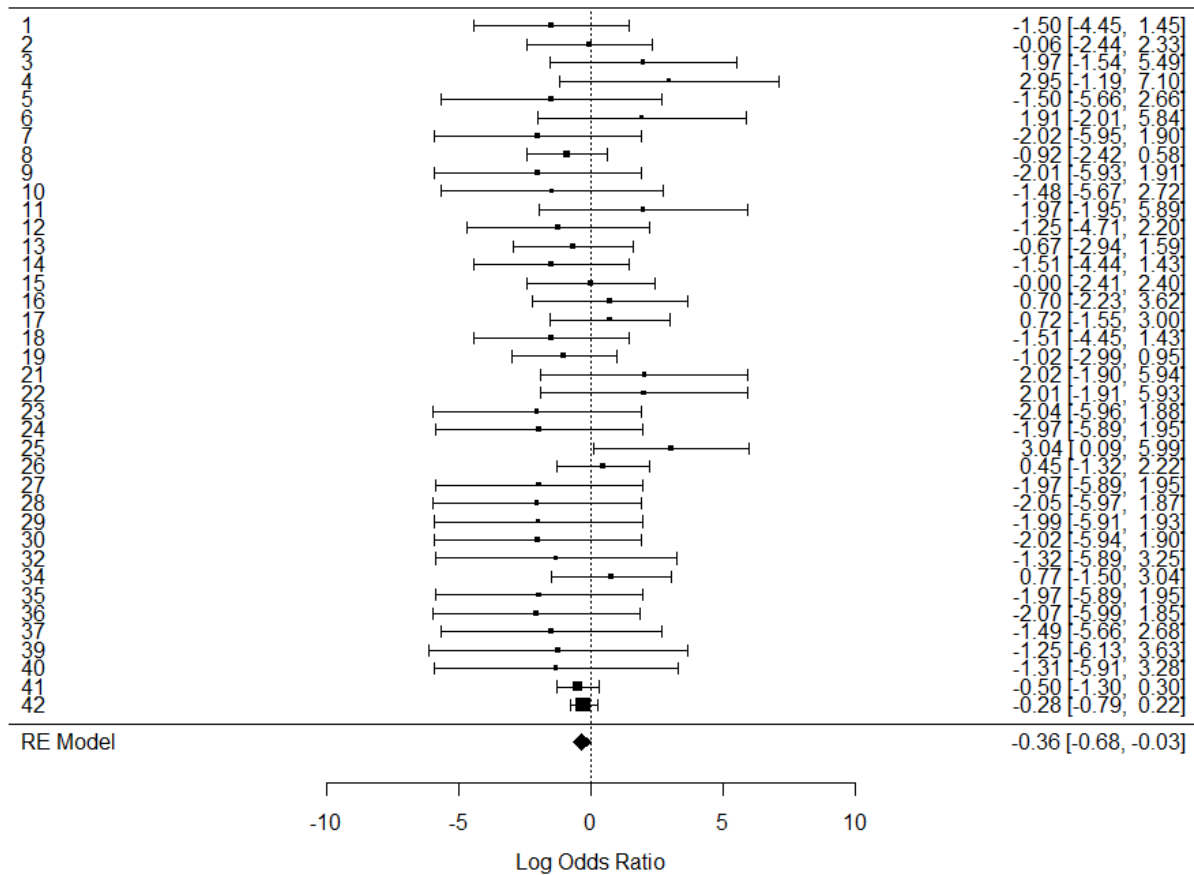|	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.1986819	0.1710201	-1.161746	0.2522240
x	0.3594419	0.1684843	2.133385	0.0390786

The p-value (highlighted in the table above) is associated with the null hypotheses that the true intercept of the regression line in the Galbraith plot is 0. Since our p-value is non-significant (0.25), we have no evidence that the true intercept is not 0. That indicates that there is no evidence of publication bias.

5. Repeat above analysis with the Peto Method.

   a. $\widehat{\theta}_i = \log\left(\frac{\left(b - \frac{(a+b)(b+d)}{(n)}\right)}{\frac{(a+b)(b+d)(a+c)(c+d)}{n^2(n-1)}}\right)$, where $n = a + b + c + d$.

   This is the percent increase (or decrease) in odds of event $Y = 1$ for Treatment group compared with control group.

   b.

|	
1	-1.50 [-4.45, 1.45]
2	-0.06 [-2.44, 2.33]
3	1.97 [-1.54, 5.49]
4	2.95 [-1.19, 7.10]
5	-1.50 [-5.66, 2.66]
6	1.91 [-2.01, 5.84]
7	-2.02 [-5.95, 1.90]
8	-0.92 [-2.42, 0.58]
9	-2.01 [-5.93, 1.91]
10	-1.48 [-5.67, 2.72]
11	1.97 [-1.95, 5.89]
12	-1.25 [-4.71, 2.20]
13	-0.67 [-2.94, 1.59]
14	-1.51 [-4.44, 1.43]
15	-0.00 [-2.41, 2.40]
16	0.70 [-2.23, 3.62]
17	0.72 [-1.55, 3.00]
18	-1.51 [-4.45, 1.43]
19	-1.02 [-2.99, 0.95]
21	2.02 [-1.90, 5.94]
22	2.01 [-1.91, 5.93]
23	-2.04 [-5.96, 1.88]
24	-1.97 [-5.89, 1.95]
25	3.04 [0.09, 5.99]
26	0.45 [-1.32, 2.22]
27	-1.97 [-5.89, 1.95]
28	-2.05 [-5.97, 1.87]
29	-1.99 [-5.91, 1.93]
30	-2.02 [-5.94, 1.90]
32	-1.32 [-5.89, 3.25]
34	0.77 [-1.50, 3.04]
35	-1.97 [-5.89, 1.95]
36	-2.07 [-5.99, 1.85]
37	-1.49 [-5.66, 2.68]
39	-1.25 [-6.13, 3.63]
40	-1.31 [-5.91, 3.28]
41	-0.50 [-1.30, 0.30]
42	-0.28 [-0.79, 0.22]
RE Model	-0.36 [-0.68, -0.03]

In the forest plot above, the diamond at the bottom of the plot (displaying the overall log odds ratio) does not quite capture 0. This indicates a significant result, implying that there is a difference between odds of a myocardial infarction between the two treatment groups.

c. $\widehat{OR} = e^{-0.3565} = 0.7001$

Rosiglitazone significantly decreases the odds of not having a myocardial infarction by about 30% compared to the control treatment. This is statistically significant (p = 0.0321).

Publication Bias investigation (may have been unnecessary – not asked for explicitly in assignment).
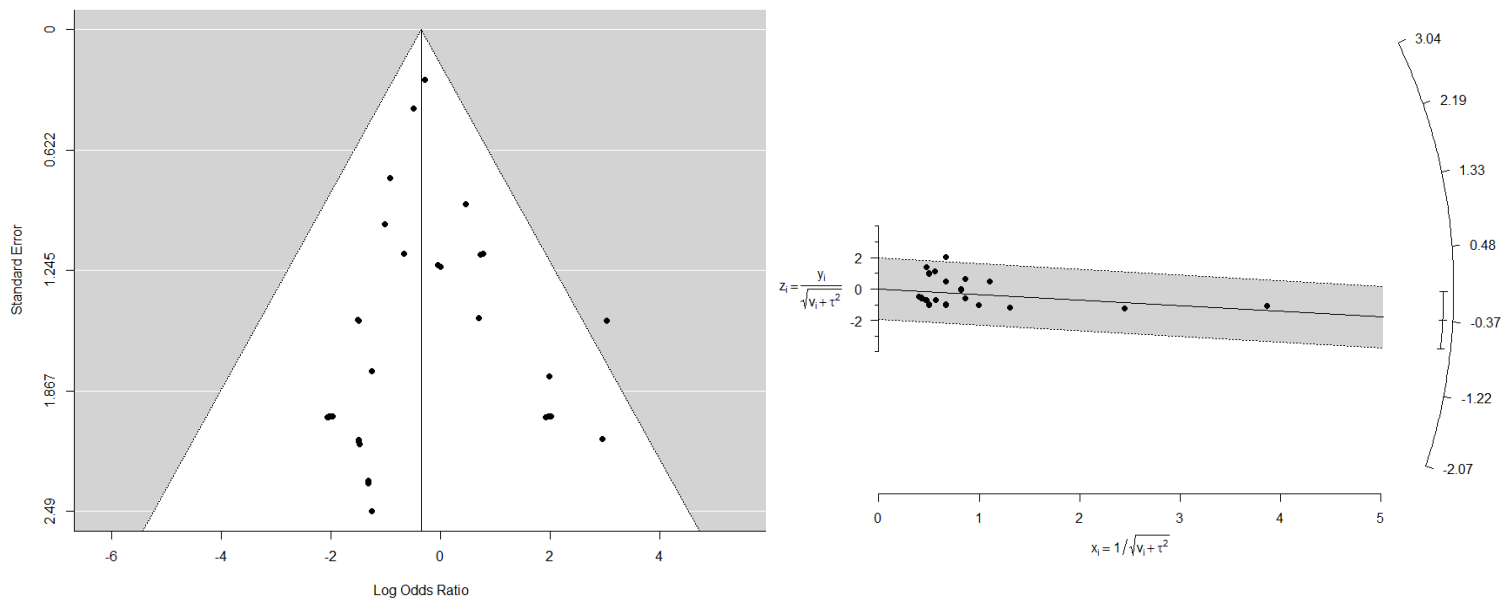
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.1140072	0.2288151	-0.4982503	0.6213381
x	-0.2664765	0.2346239	-1.1357605	0.2635617

The funnel plot is definitely not symmetric. There is a distinct group of points going down the right hand side of the funnel. If the other group of points was not there, I think that we would definitely have a publication bias. However, I don't think we can say anything conclusive from the funnel plot alone. The radial plot gives a little more evidence that we do not have a publication bias. In addition, the regression line in the radial plot seems to have its intercept going right through the origin. The numeric diagnostic confirms this – with a p-value of 0.62, we have absolutely no evidence that the line in the radial plot does not go through the origin.

In summary, with all these diagnostics considered, I don't think that we have any strong evidence of a publication bias.

    d. How does your result in Exercise 5c compare with the findings reported in the Results paragraph of the Abstract of the NEJM article?
My p-value was identical to theirs. It looks like they defined their 'event' differently than I did. They calculated their odds ratio in terms of odds for myocardial infarction, while I calculated mine in terms of odds of *no* myocardial infarction.

6. The Peto method shows a significant rosiglitazone effect on the myocardial infarction rate (5c above), but the "+0.5" method does not (4c above). The figures below summarize the odds ratio estimates from each of the studies using these two methods. There are ten studies with Peto odds ratio estimates greater than 5 (highlighted in left plot). In each of these ten studies, there were no myocardial infarction events in the control group, and just one myocardial infarction event in the rosiglitazone group.
   a. The Peto method would drop studies from the meta-analysis if they had no myocardial infarctions in either group. It would drop them because it presents problems mathematically when we try to calculate the odds ratio.
   b. In this case, we obtain an effect size of 0.2336. This translates to an odds ratio of: $\widehat{OR} = e^{0.2336} = 1.2631$. This indicates that the odds of not having a myocardial infarction increase by about 26% for the Rosiglitazone group as compared with the control group. However, this difference is not statistically significant (p = 0.1755).

R Code:

```
# BioStat Homework 7

# Exercise 4
library(metafor)
data <- read.csv("http://www.stat.usu.edu/jrstevens/biostat/data/rosiglitazone.csv")

Fishers.p <- rep(NA,nrow(data))

for(i in 1:nrow(data))
{
  mat <- matrix(as.numeric(data[i,]), ncol = 2)
  Fishers.p[i] <- fisher.test(mat)$p.value
}

summary(data)
data <- data.frame(data)
#round(cbind(data,as.numeric(Fishers.p)),5)
data <- cbind(data, Fishers.p)

a <- data$RosMI # Ros, MI
b <- data$RosNum - data$RosMI # Ros, no MI
c <- data$CtlMI # Ctl, MI
d <- data$CtlNum - data$CtlMI # Ctl, no MI
# theta.hat <- log( (b + 0.5) * (c + 0.5) / ( (a + 0.5) * (d + 0.5) ))
# var.theta.hat <- 1/(b + 0.5) + 1/(c + 0.5) + 1/(a + 0.5) + 1/(d + 0.5)
# w.RE <- 1/var.theta.hat
# theta.RE <- sum(w.RE * theta.hat) / sum(w.RE)
```

```r
# var.RE <- 1/sum(w.RE)
# z.RE <- theta.RE/sqrt(var.RE)
# p.RE <- 2 * (1 - pnorm(abs(z.RE)))
# round(c(z.RE,p.RE),5)
# OR.RE <- exp(theta.RE)
# OR.RE

result <- rma.uni(
  ai = a,
  bi = b,
  ci = c,
  di= d,
  measure = 'OR',
  add = 0.5, to = 'all',
  method = 'DL',
  slab = 1:42
)
summary(result)

forest(result)
funnel(result)
radial(result)
z <- result$yi/sqrt(result$vi + result$tau2)
x <- 1/sqrt(result$vi + result$tau2)
fit <- lm(z~x)
summary(fit)$coeff


# Peto Method #
result_peto <- rma.uni(
  ai = a,
  bi = b,
  ci = c,
  di= d,
  add = 0, to = 'all',
  measure = 'PETO',
  method = 'DL',
  slab = 1:42
)
summary(result_peto)

forest(result_peto)
funnel(result_peto)
radial(result_peto)
z <- result_peto$yi/sqrt(result_peto$vi + result_peto$tau2)
x <- 1/sqrt(result_peto$vi + result_peto$tau2)
fit <- lm(z~x)
summary(fit)$coeff
```

```r
ids.drop <- c("49653/127","49653/128","49653/136","49653/143","49653/145",
        "49653/147","49653/162","49653/284","SB-712753/002","SB-712753/003")
data6 <- data[!is.element(data$StudyID,ids.drop),]

a6 <- data6$RosMI # Ros, MI
b6 <- data6$RosNum - data6$RosMI # Ros, no MI
c6 <- data6$CtlMI # Ctl, MI
d6 <- data6$CtlNum - data6$CtlMI # Ctl, no MI

result_peto6 <- rma.uni(
  ai = a6,
  bi = b6,
  ci = c6,
  di= d6,
  add = 0, to = 'all',
  measure = 'PETO',
  method = 'DL',
  slab = 1:32
)
summary(result_peto6)

forest(result_peto6)
funnel(result_peto6)
radial(result_peto6)
z <- result_peto6$yi/sqrt(result_peto6$vi + result_peto6$tau2)
x <- 1/sqrt(result_peto6$vi + result_peto6$tau2)
fit <- lm(z~x)
summary(fit)$coeff
```