HW6

Jinhong Du, 15338039

目录

1 Problem Description

	Gentamicin		Tobramycin					
Study	No. of subjects	No. of positives[a]	No. of subjects	No. of positives[a]	Odds ratio[b]	y_i^c	w_i^d	w_i^{*e}
1. Walker	40	7	40	2	0.25	-1.394	1.430	1.191
2. Wade	43	13	47	11	0.71	-0.349	4.367	2.709
3. Greene	11	2	15	2	0.69	-0.368	0.842	0.753
4. Smith	72	19	74	9	0.39	-0.951	5.051	2.957
5. Fong	102	18	103	15	0.80	-0.229	6.873	3.500
6. Brown	103	5	96	2	0.42	-0.875	1.387	1.161
7. Feig	25	10	29	8	0.57	-0.560	2.947	2.086
8. Matzke	99	9	97	17	2.13	+0.754	5.167	2.996

[a] Number who developed nephrotoxicity.

[b] Odds in favor of nephrotoxicity for tobramycin patients / odds in favor of nephrotoxicity for gentamicin patients.

[c] $y_i = \ln(\hat{OR}_i)$

[d] $w_i = \dfrac{1}{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}}$

[e] $w_i^* = \dfrac{1}{\frac{1}{w_i} + \hat{\Delta}^2}$

Suppose there is an underlying log odds ratio θ_i for the ith study, which is estimated by $y_i = \ln(\hat{OR}_i)$ $i = 1, \ldots, 8$, where the estimated OR_i are given in Table 13.31 in the Odds ratio column. We assume there is within-study variation of y_i about θ_i , where the variance of y_i is

$$s_i^2 = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i} = \frac{1}{w_i}$$

and a_i, b_i, c_i, and d_i are the cell counts in the 2×2 table for the ith study.

2 Random-effect Model

We also assume that there is between-study variation of Δ_i about an average true $\log OR$ μ over all studies so that

$$\theta_i = \mu + \delta_i$$

and

$$Var(\Delta_i) = \Delta^2$$

To estimate μ, we calculate a weighted average of the study-specific $\log OR$s given by

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i^* y_i}{\sum_{i=1}^{k} w_i^*}$$

where $w_i^* = \dfrac{1}{s_i^2 + \hat{\Delta}^2}$, i.e., the weight for the ith study is inversely proportional to the total variance for that study (which equals $s_i^2 + \Delta^2$ where s_i^2 =within-study variance in the ith group), and

$$se(\hat{\mu}) = \frac{1}{\sqrt{\sum_{i=1}^{k} w_i^*}}$$

It can be shown that the best estimate of Δ^2 is given by

$$\hat{\Delta}^2 = \max\left\{0, [Q_w - (k-1)]\left(\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}\right)^{-1}\right\}$$

where

$$Q_w = \sum_{i=1}^{k} w_i (y_i - \overline{y}_w)^2$$

and

$$\overline{y}_w = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i}$$

To test

$$H_0 : OR = 1 \qquad H_1 : OR \neq 1$$

is equivalent to test

$$H_0 : \mu = 0 \qquad H_1 : \mu \neq 0$$

The test statistic is given by

$$\frac{\hat{\mu}}{se(\hat{\mu})} \overset{H_0}{\sim} N(0,1)$$

```
k <- 8
table_data <- matrix(
    c(40, 7 , 40, 2 , 0.25, -1.394, 1.430, 1.191,
      43, 13, 47, 11, 0.71, -0.349, 4.367, 2.709,
      11, 2 , 15, 2 , 0.69, -0.368, 0.842, 0.753,
      72, 19, 74, 9 , 0.39, -0.951, 5.051, 2.957,
      102,18, 103,15, 0.80, -0.229, 6.873, 3.500,
      103,5 , 96, 2 , 0.42, -0.875, 1.387, 1.161,
      25, 10, 29, 8 , 0.57, -0.560, 2.947, 2.086,
```

```
      99, 9 , 97, 17, 2.13, +0.754, 5.167, 2.996), nrow = k, ncol = 8, byrow = T)

y <- table_data[,6]
w <- table_data[,7]
w_star <- table_data[,8]

y_bar_w <- as.vector(crossprod(w,y) / sum(w))
Qw <- sum(w * (y - y_bar_w)^2)
Delta2 <- max(c(0,(Qw - (k-1))/ (sum(w) - sum(w^2)/sum(w))))
se_mu <- 1/ sqrt(sum(w_star))
mu <- as.vector(crossprod(w_star, y) / sum(w_star))
z <- mu / se_mu
cat('The p-value is ', 2 * (1 - pnorm(abs(z))),'\n')
cat('The point estimate of the overall OR is ',exp(mu),'\n')
cat('The 95% confidence interval for ',exp(mu),
    ' is (',exp(mu - qnorm(0.975)*se_mu),',',exp(mu + qnorm(0.975)*se_mu),')\n')
```

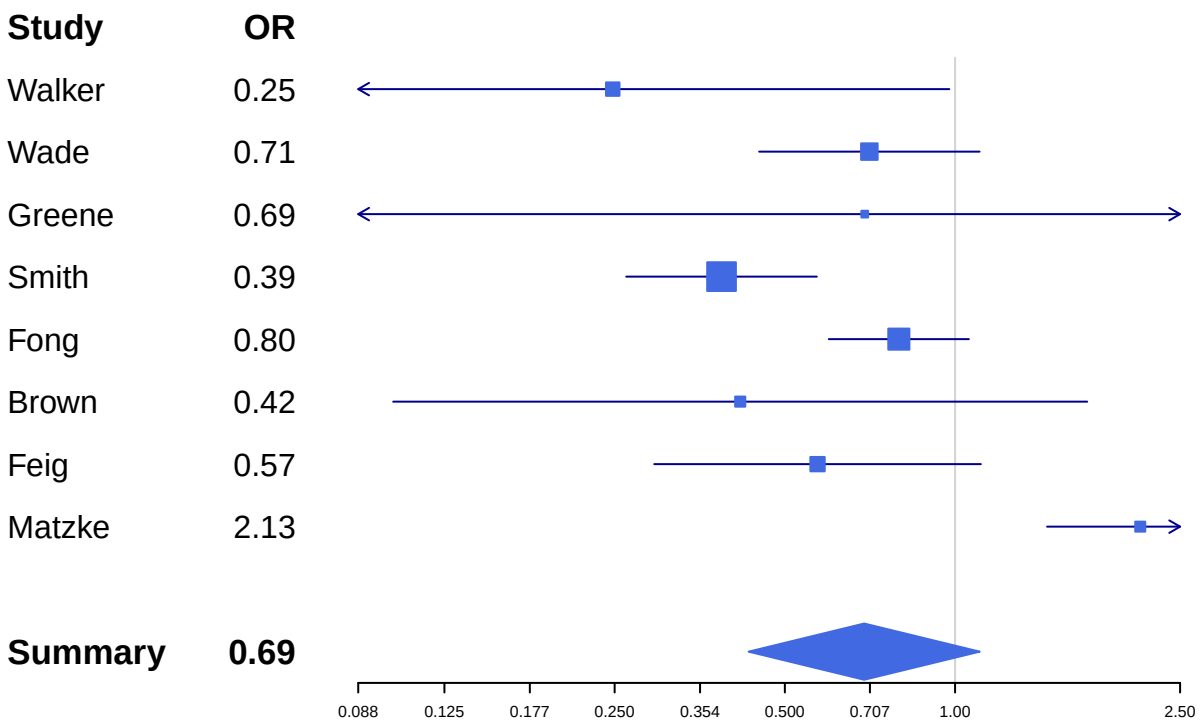
```
## The p-value is  0.1231928
## The point estimate of the overall OR is  0.6907006
## The 95% confidence interval for  0.6907006  is ( 0.4314731 , 1.105671 )
```

Since the p-value is bigger than $\alpha = 0.05$, we do not reject H_0. Hence the true OR does not significantly differ from 1.

```
library(forestplot)
cochrane_from_rmeta <-
  structure(list(
    mean  = c(NA, NA, exp(y), NA, exp(mu)),
    lower = c(NA, NA, exp(y - qnorm(0.975)/w), NA, exp(mu - qnorm(0.975)*se_mu)),
    upper = c(NA, NA, exp(y + qnorm(0.975)/w), NA, exp(mu + qnorm(0.975)*se_mu))),
    .Names = c("mean", "lower", "upper"),
    row.names = c(NA, -12L),
    class = "data.frame")

tabletext<-cbind(
  c("", "Study", "Walker", "Wade",
    "Greene", "Smith", "Fong", "Brown",
    "Feig", "Matzke", NA, "Summary"),
  c("", "OR", "0.25","0.71","0.69","0.39","0.80","0.42","0.57","2.13", NA, "0.69"))
```

```
forestplot(tabletext,
           cochrane_from_rmeta,
           is.summary=c(TRUE,TRUE,rep(FALSE,9),TRUE),
           clip=c(0.1,2.5),
           xlog=T,
           col=fpColors(box="royalblue",line="darkblue", summary="royalblue"))
```



3 Test of Homogeneity of Odds Ratios

To test the hypothesis

$$H_0 : \theta_1 = \cdots = \theta_k \qquad H_1 : \exists\ i \neq j,\ \theta_i \neq \theta_j$$

The statistic is given by $Q_w = \sum_{i=1}^{k} w_i (y_i - \overline{y}_w)^2 \overset{H_0}{\sim} \chi^2_{k-1}$

```
cat('The p-value is ',1 - pchisq(Qw, k-1))
```

```
## The p-value is  0.173318
```

Since the p-value is bigger than $\alpha = 0.05$, we cannot reject H_0, i.e., there is not significant heterogeneity among the study-specific ORs.

4 Fixed-effect Model

In fixed-effect model, the estimated variances in all studies are given by w_i $(i = 1, 2, \cdots, 8)$ and the estimated variance of weighted average $\log OR$ is given by

$$se(\hat{\mu}) = \frac{1}{\sqrt{\sum_{i=1}^{k} w_i}}$$

```
y <- table_data[,6]
w <- table_data[,7]
se_mu <- 1/ sqrt(sum(w))
mu <- as.vector(crossprod(w, y) / sum(w))
z <- mu / se_mu
cat('The p-value is ', 2 * (1 - pnorm(abs(z))),'\n')
cat('The point estimate of the overall OR is ',exp(mu),'\n')
cat('The 95% confidence interval for ',exp(mu),
    ' is (',exp(mu - qnorm(0.975)*se_mu),',',exp(mu + qnorm(0.975)*se_mu),')\n')
```

```
## The p-value is  0.08335919
## The point estimate of the overall OR is  0.7211897
## The 95% confidence interval for  0.7211897  is ( 0.4981626 , 1.044066 )
```

```
library(forestplot)
cochrane_from_rmeta <-
  structure(list(
    mean  = c(NA, NA, exp(y), NA, exp(mu)),
    lower = c(NA, NA, exp(y - qnorm(0.975)/w), NA, exp(mu - qnorm(0.975)*se_mu)),
    upper = c(NA, NA, exp(y + qnorm(0.975)/w), NA, exp(mu + qnorm(0.975)*se_mu))),
    .Names = c("mean", "lower", "upper"),
    row.names = c(NA, -12L),
    class = "data.frame")

tabletext<-cbind(
  c("", "Study", "Walker", "Wade",
    "Greene", "Smith", "Fong", "Brown",
    "Feig", "Matzke", NA, "Summary"),
  c("", "OR", "0.25","0.71","0.69","0.39","0.80","0.42","0.57","2.13", NA, "0.72"))

forestplot(tabletext,
           cochrane_from_rmeta,
           is.summary=c(TRUE,TRUE,rep(FALSE,9),TRUE),
           clip=c(0.1,2.5),
```

```
        xlog=T,
        col=fpColors(box="royalblue",line="darkblue", summary="royalblue"))
```