ph1861\_hw2\_ygu5

Yue Gu

2023-11-21

#Meta-Analysis Problem #1 (50 points) The following paper reviewed and meta-analyzed the proportion of Campylobacter cases that develop chronic sequelae, based on studies published prior to July 2011. (Keithlin 2014, “Systematic review and meta-analysis of the proportion of Campylobacter cases that develop chronic sequelae”). Import the number of cases that developed reactive arthritis (ReA) from page 7 of the article and follow the steps below to conduct a meta-analysis of the proportion/prevalence.

Please input data in the exact order from the paper. Calculate the proportion of people with Campylobacter who developed a ReA for each study. Add a count of 0.5 both to those reporting ReA outcome of 0% and to those totals. Calculate the logit of the outcome and logit of the standard error using the equation from Lipsey & Wilson. [Eq 3.5, p40].

# read data  
keith\_data = read.csv("./data/Keithlin2014\_ReA.csv") %>%   
 janitor::clean\_names() %>%   
 mutate(prop\_seq = round(number\_developing\_sequelae/number\_of\_people\_with\_campylobacter,4)) %>%   
 mutate(number\_of\_people\_with\_campylobacter\_add = ifelse(prop\_seq==0.0000, number\_of\_people\_with\_campylobacter+0.5, number\_of\_people\_with\_campylobacter),  
 number\_developing\_sequelae\_add = ifelse(prop\_seq==0.0000, number\_developing\_sequelae+0.5, number\_developing\_sequelae)) %>%   
 mutate(prop\_seq\_add = number\_developing\_sequelae\_add/number\_of\_people\_with\_campylobacter\_add,  
 logit\_outcome = log(prop\_seq/(1-prop\_seq)),  
 se\_outcome = sqrt(1/(number\_of\_people\_with\_campylobacter\_add\*prop\_seq\_add)+1/(number\_of\_people\_with\_campylobacter\_add\*(1-prop\_seq\_add))))  
  
# check total number of cases  
s1=sum(keith\_data$number\_developing\_sequelae\_add)  
# check total number  
s2=sum(keith\_data$number\_of\_people\_with\_campylobacter\_add)  
# check the proportion developing ReA  
p1 = s1/s2

Now that the data are set, answer following questions with corresponding STATA or R code:  
(a) List the number of cases, total number, the proportion developing ReA, the logit outcome, and the logit standard error for the K=25 studies.

# the logit outcome for the K=25 studies  
knitr::kable(keith\_data %>% select(first\_author\_year\_reference\_number, logit\_outcome))

| first\_author\_year\_reference\_number | logit\_outcome |
| --- | --- |
| Ternhag, [52] | -8.111428 |
| Schoenberg-Norio, jejuni [51] | -3.183275 |
| Townes, [47] | -4.269191 |
| Kosunen, [39] | -3.731341 |
| Petersen, [27] | -Inf |
| Short, [25] | -Inf |
| Hannu, [43] | -2.528269 |
| Hannu, [43] | -2.371141 |
| Hannu, [43] | -2.599044 |
| Hannu, [43] | -1.891268 |
| Hannu, [43] | -Inf |
| Melby, [32] | -5.060886 |
| Pitkanen, [23] | -2.871116 |
| Locht, [37] | -1.687537 |
| Schiellerup, [41] | -1.895664 |
| Hannu, [54] | -3.635228 |
| Pitkanen, [29] | -2.989555 |
| Helms, [38] | -6.724233 |
| Ponka, [40] | -3.832326 |
| Doorduyn, [44] | -3.029746 |
| Eastmond, [35] | -4.462676 |
| Eastmond [35] | -Inf |
| Melby, [55] | -3.623315 |
| Gumpel, [49] | -1.139566 |
| Bremell, [33] | -4.171143 |

# the standard error for the K=25 studies  
knitr::kable(keith\_data %>% select(first\_author\_year\_reference\_number, se\_outcome))

| first\_author\_year\_reference\_number | se\_outcome |
| --- | --- |
| Ternhag, [52] | 0.2582326 |
| Schoenberg-Norio, jejuni [51] | 0.3608065 |
| Townes, [47] | 0.1752951 |
| Kosunen, [39] | 0.3577625 |
| Petersen, [27] | 1.4228107 |
| Short, [25] | 1.4375906 |
| Hannu, [43] | 0.1549041 |
| Hannu, [43] | 0.1450038 |
| Hannu, [43] | 0.1703968 |
| Hannu, [43] | 0.3792993 |
| Hannu, [43] | 1.4411534 |
| Melby, [32] | 1.0031596 |
| Pitkanen, [23] | 0.5934655 |
| Locht, [37] | 0.2094907 |
| Schiellerup, [41] | 0.0937037 |
| Hannu, [54] | 0.3377035 |
| Pitkanen, [29] | 0.3416105 |
| Helms, [38] | 0.2133312 |
| Ponka, [40] | 0.4126461 |
| Doorduyn, [44] | 0.2289442 |
| Eastmond, [35] | 1.0057307 |
| Eastmond [35] | 1.4226066 |
| Melby, [55] | 0.7164728 |
| Gumpel, [49] | 0.4062019 |
| Bremell, [33] | 1.0076629 |

The number of cases for the K=25 studies is 448, total number is 8.318^{4}, the proportion developing ReA is 0.0053859, the logit outcome and standard error for each studies could be found above  
(b) Run a fixed effects meta-analysis using the exponential form (ie, exponentiating the result back to the original scale). Attach the forest plot.

# run fixed effects meta-analysis for a single proportion  
library(meta)  
m.bin <- metaprop(number\_developing\_sequelae\_add,number\_of\_people\_with\_campylobacter\_add,  
 data = keith\_data,  
 studlab = paste(first\_author\_year\_reference\_number),  
 comb.fixed = T, comb.random = T,  
 method = 'GLMM',sm = "PLOGIT",  
 model.glmm = "UM.FS")

## Warning: Normal approximation confidence interval (argument method.ci = "NAsm") used as  
## at least one number of events contains a non-integer value.

## Warning in eval(family$initialize): non-integer counts in a binomial glm!

## Warning in eval(family$initialize, rho): non-integer counts in a binomial glm!

m.bin

## Number of studies: k = 25  
## Number of observations: o = 83180  
## Number of events: e = 448  
##   
## proportion 95%-CI  
## Common effect model 0.0054 [0.0049; 0.0059]  
## Random effects model 0.0254 [0.0132; 0.0482]  
##   
## Quantifying heterogeneity:  
## tau^2 = 2.4692; tau = 1.5714; I^2 = 97.7% [97.2%; 98.1%]; H = 6.55 [5.96; 7.21]  
##   
## Test of heterogeneity:  
## Q d.f. p-value  
## Wald 1030.79 24 < 0.0001  
## LRT 2033.29 24 0  
##   
## Details on meta-analytical method:  
## - Random intercept logistic regression model  
## - Maximum-likelihood estimator for tau^2  
## - Logit transformation

# forest plot  
forest(m.bin, leftcols = c('studlab'), overall = T, weight.study="random",hetstat = T, overall.hetstat = T)

A graph with numbers and a line

Description automatically generated with medium confidence

The results from fixed effects meta-analysis and forest plot are shown above.  
(c) What is the pooled estimate? Is it significant? Please interpret the result.  
The pooled estimate for Heterogeniety across all the studies is 0.01 for common-effects model and 0.03 for random effects model and with p<0.01, we interpret it as 98% of the variation across the studies is due to heterogeneity rather than chance and 2.4692 is the amount of heterogeneity between the study effect esimates; with significant p-value, we conclude that there is significant heterogeneity between the 25 studies, and further random effect meta-analysis is needed.  
(d) If necessary, run a random effect meta-analysis using exponential form.

# random effect meta-analysis (same command in R)  
m.bin <- metaprop(number\_developing\_sequelae\_add,number\_of\_people\_with\_campylobacter\_add,  
 data = keith\_data,  
 studlab = paste(first\_author\_year\_reference\_number),  
 comb.fixed = T, comb.random = T,  
 method = 'GLMM',sm = "PLOGIT",  
 model.glmm = "UM.FS")

## Warning: Normal approximation confidence interval (argument method.ci = "NAsm") used as  
## at least one number of events contains a non-integer value.

## Warning in eval(family$initialize): non-integer counts in a binomial glm!

## Warning in eval(family$initialize, rho): non-integer counts in a binomial glm!

m.bin

## Number of studies: k = 25  
## Number of observations: o = 83180  
## Number of events: e = 448  
##   
## proportion 95%-CI  
## Common effect model 0.0054 [0.0049; 0.0059]  
## Random effects model 0.0254 [0.0132; 0.0482]  
##   
## Quantifying heterogeneity:  
## tau^2 = 2.4692; tau = 1.5714; I^2 = 97.7% [97.2%; 98.1%]; H = 6.55 [5.96; 7.21]  
##   
## Test of heterogeneity:  
## Q d.f. p-value  
## Wald 1030.79 24 < 0.0001  
## LRT 2033.29 24 0  
##   
## Details on meta-analytical method:  
## - Random intercept logistic regression model  
## - Maximum-likelihood estimator for tau^2  
## - Logit transformation

Based on the results, the pooled estimates for random effects is 0.0254 with wider CI of [0.0132, 0.0482], which are larger proportion and wider CI compared to common effects proportion.  
(e) Why don’t we have to convert the proportion to Cohen’s D?  
The Cohen’s D need assumptions of normal distribution and we don’t necessarily have norm-distribution here across all the studies. Also since we found significant heterogeneity across studies, the Cohen’s D are not appropriate to use.  
Write a short methods and results paragraph as if you were reporting these two sections in a journal article (refer to our in-class case studies).  
Methods: We collected 25 studies measuring Campylobacter and chronic sequelae with ReA outcomes, the number of people with Campylobacter and the number developing sequelae are mainly used to calculate the percentage of outcome. 0.5 are added to the total number of people with Campylobacter for the outcomes that have 0% keep 4 decimal points in the original scales. Based on the collected studies, we run the meta-analysis investigating both random-effects and fixed-effects, and forest plot was used to demonstrate the weight of each studies and confidence interval for each proportion of ReA outcomes.  
Results: Of the 25 studies included for meta-analysis, with the addition of 0.5, the total number of cases is 448, total number of subjects included is 83180 with the overall outcome of developing ReA as 0.54%. Results from meta-anlysis showed significant heterogeineity from the studies () and a random-effects MA is needed. Based on the results, the pooled proportion for ReA outcome is 0.0254 with 95% CI of [0.0132,0.0482] and forest plot supports the highest weights is contributed from Gumpel et al. study with 0.24 proportion.

#Meta-Analysis Problem #2 (50 points) We conducted a systematic review (SR) of second generation antidepressants that compared fluoxetine to any other anti-depressant. Although this SR resulted in an extensive network of multiple treatment comparisons, for this assignment, we will treat this review as “pairwise.” Thus we will assume all comparator arms to fluoxetine are called “other active treatment.” The Excel file is attached above. The outcome is “efficacy” and the number experiencing efficacy and total number are given. If a study has 3 treatments, you will only compare fluoxetine with the first treatment given in that study. (a) Conduct an appropriate pairwise meta-analysis of these RCT data. Present the forest plot and pooled effect size.

# read data  
fluo\_data = readxl::read\_xlsx("./data/HW2 Fluoxetine RCT Data -1.xlsx",sheet = 1) %>%   
 janitor::clean\_names()  
# reshape the data into binary  
fluo\_data\_wide = fluo\_data %>%   
 mutate(study\_id = ifelse(is.na(study\_id), lag(study\_id), study\_id)) %>%   
 filter(study\_id!="") %>%   
 group\_by(study\_id) %>%  
 mutate(row = row\_number(),  
 trt = if\_else(treatment=="fluoxetine", "fluoxetine", "others")) %>%  
 ungroup() %>%  
 pivot\_wider(names\_from = trt, values\_from = c(number\_efficacy,total\_n), names\_sep = "\_") %>%  
 select(-row,-treatment,-weeks\_followup,-outcome\_scale) %>%   
 mutate(number\_efficacy\_others = as.numeric(number\_efficacy\_others),  
 number\_efficacy\_fluoxetine = as.numeric(number\_efficacy\_fluoxetine)) %>%   
 group\_by(study\_id) %>%  
 summarise(across(everything(), ~ if(all(is.na(.))) NA else sum(., na.rm = T)))

## Warning: There were 2 warnings in `mutate()`.  
## The first warning was:  
## ℹ In argument: `number\_efficacy\_others = as.numeric(number\_efficacy\_others)`.  
## Caused by warning:  
## ! NAs introduced by coercion  
## ℹ Run ]8;;ide:run:dplyr::last\_dplyr\_warnings()dplyr::last\_dplyr\_warnings()]8;; to see the 1 remaining warning.

# keep only the trt and control data  
fluo\_data\_wide$flu\_e1=fluo\_data\_wide$number\_efficacy\_fluoxetine  
fluo\_data\_wide$flu\_e0=fluo\_data\_wide$total\_n\_fluoxetine - fluo\_data\_wide$number\_efficacy\_fluoxetine  
fluo\_data\_wide$others\_e1=fluo\_data\_wide$number\_efficacy\_others  
fluo\_data\_wide$others\_e0=fluo\_data\_wide$total\_n\_others - fluo\_data\_wide$number\_efficacy\_others  
fluo\_data\_wide[is.na(fluo\_data\_wide)]=.5  
# check data  
new\_list = fluo\_data\_wide[,c("study\_id", "flu\_e0", "flu\_e1", "others\_e0", "others\_e1")]  
new\_list

## # A tibble: 53 × 5  
## study\_id flu\_e0 flu\_e1 others\_e0 others\_e1  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 2906/421 0.5 0.5 0.5 0.5  
## 2 29060/365 38 32 35 33   
## 3 Aguglia1993 30 26 17 35   
## 4 Alves1999 12 35 4 36   
## 5 Amini2005 10 8 6 12   
## 6 Annseau1994 50 43 65 32   
## 7 Bennie1995 81 63 69 73   
## 8 Bosc1997a 72 55 77 49   
## 9 Bosc1997b 38 51 33 46   
## 10 Bougerol1997a 71 113 58 115   
## # ℹ 43 more rows

# conduct pairwise meta-analysis  
m.bin <- metabin(others\_e1, total\_n\_others, flu\_e1, total\_n\_fluoxetine,  
 data = fluo\_data\_wide,  
 studlab = paste(study\_id),  
 comb.fixed = T,comb.random = T,  
 method = 'mh',sm = "OR",  
 model.glmm = "UM.FS")  
m.bin

## Number of studies: k = 53  
## Number of observations: o = 10386  
## Number of events: e = 5526  
##   
## OR 95%-CI z p-value  
## Common effect model 1.1047 [1.0179; 1.1989] 2.38 0.0171  
## Random effects model 1.1350 [1.0275; 1.2537] 2.50 0.0126  
##   
## Quantifying heterogeneity:  
## tau^2 = 0.0305 [0.0000; 0.1011]; tau = 0.1747 [0.0000; 0.3180]  
## I^2 = 23.3% [0.0%; 45.8%]; H = 1.14 [1.00; 1.36]  
##   
## Test of heterogeneity:  
## Q d.f. p-value  
## 67.81 52 0.0694  
##   
## Details on meta-analytical method:  
## - Mantel-Haenszel method  
## - Inverse variance method  
## - Restricted maximum-likelihood estimator for tau^2  
## - Q-Profile method for confidence interval of tau^2 and tau

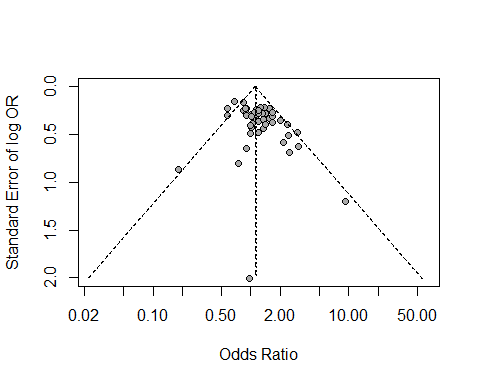
# forest plot  
forest(m.bin, leftcols = c('studlab'), cex=0.5)

A screenshot of a computer screen

Description automatically generated

The pooled effect size are shown in the forest plot with 1.1 OR from common effect model and 1.14 from random effects model.  
(b) Does there appear to be publication bias? Please justify in 2 sentences or less.

# plot funnel plot  
funnel.meta(m.bin, ylab = "Standard Error of log OR")



# egger's test  
metabias(m.bin, method = "Egger")

## Linear regression test of funnel plot asymmetry  
##   
## Test result: t = 2.74, df = 51, p-value = 0.0085  
##   
## Sample estimates:  
## bias se.bias intercept se.intercept  
## 1.0872 0.3971 -0.2111 0.1218  
##   
## Details:  
## - multiplicative residual heterogeneity variance (tau^2 = 1.1593)  
## - predictor: standard error  
## - weight: inverse variance  
## - reference: Egger et al. (1997), BMJ

Based on the funnel plot, we see a slight asymmetric pattern for the standard error of log OR and the p-value=0.0085 from the Egger’s test is significant, indicating the publication bias might be concerning and is significantly strong.  
(c) Is a random or fixed effects model more appropriate here? Justify in 2 sentences or less. Common effects model is more appropriate since we didn’t found a significant heterogeneity across the studies from the meta-analysis model and .  
(d) Write a very short Methods and a Results paragraph summarizing your findings from (a-c).  
53 studies are included for meeting the inclusion criteria and we treat fluoxetine as control, and all other treatments as active treatments. The meta-analysis results are shown in the forest plot that we observe the strongest weight from the studies are Winokur2000 and the smallest weight study is Taner2006, while other studies mostly fall near the pooled OR line. We further check whether there is publication bias from the studies, from the funnel plot, we observed a significantly strong PB and Egger’s test generates a significant p-value, confirming there is significant sub-group effects. In conclusion, given the non-significant heterogeneity across the studies, a common effects model would be better to consider with 1.10 as the pooled OR and since there is no PB found, subgroup analysis might be considered to further conduct. (e) Would you consider any subgroup or sensitivity analysis for this meta-analysis? Explain which one(s) and justify your answer.  
Yes, since we did observe some asymmetric patterns from the funnels plot and the Egger’s test is significant, there are significant small study effects in the study and futher subgroup or sensitivity analysis is needed to find which subgroup effects lead to the PB.