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_campyl
         number_developing_sequelae_add = ifelse(prop_seq==0.0000, number_developing_sequelae+0.5, numbe
  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_

# 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

first_author_year_reference_number	logit_outcome
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

first_author_year_reference_number	se_outcome
Bremell, [33]	1.0076629

The number of cases for the K=25 studies is 448, total number is 8.318×10^4, the proportion developing ReA is 0.0053859, the logit outcome and standard error for each studies could be found below.
(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.
```

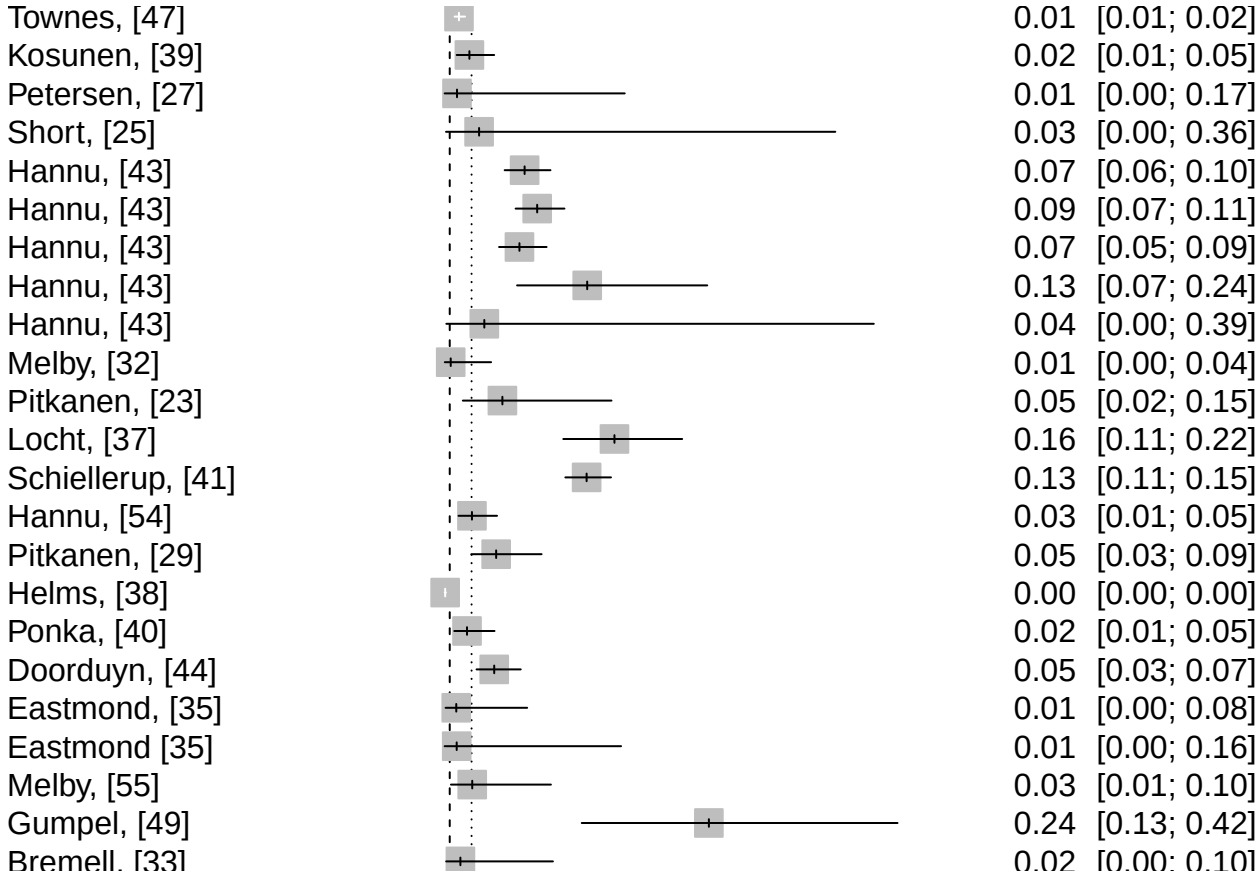
```
## Warning in eval(family$initialize): non-integer counts in a binomial glm!
```

```
## 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'),weight.study = "common")
```



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 $I^2 = 0.98, \tau^2 = 2.4692$ 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: