SST508 Practical 2: Comparison of Rates Notebook

Course Code: *SST508*

Course Name: *Survival Analysis*

Student Registration Number: *I56/32988/2019*

QUESTION 1:

1a):

In a randomised controlled trial of a malaria vaccine, antibodies to the plasmodium CS protein, an indicator to past exposure to malaria, were measured at enrolment. Use the data in the table below to decide whether baseline CS antibody status is associated with malaria incidence during follow-up.

Baseline anti-CS antibodies concentration	D (Number of Malaria cases)	Y (months at risk)	Rate
$<$ median	84	272.2	0.31
$\geq$ median	71	335.8	0.21

Solution for 1a):

The rate ratio comparing the exposed (the group with low antibody concentration) with the unexposed (the group with high antibody concentration) will be as follows:

$$RR = \frac{D_1 * Y_0}{D_0 * Y_1} = \frac{84 * 335.8}{71 * 272.2} = 1.4595 = 1.46$$

Now that the Rate ratio has been calculated, a two-tailed hypothesis test was conducted to determine if baseline CS antibody status is associated with malaria incidence during follow-up.

$$H_0 : RR = 1$$

$$H_1 : RR \neq 1$$

$$\therefore$$

$$H_0 : log(RR) = 0$$

$$H_1 : log(RR) \neq 0$$

Let the level of significance be:

$$\alpha = 0.05$$

Let **z** *be the test statistic*

$$z_{cal} = \frac{log(RR_{H1}) - log(RR_{H0})}{s.e._{log(RR_{H1})}}$$

$$z_{cal} = \frac{log(RR_{H1}) - log(1)}{s.e._{log(RR_{H1})}}$$

$$z_{cal} = \frac{log(RR_{H1}) - 0}{s.e._{log(RR_{H1})}} = \frac{log(RR_{H1})}{s.e._{log(RR_{H1})}}$$

$$log(RR_{H1}) = log(RR) = log(1.46) = 0.3784$$

$$s.e._{log(RR_{H1})} = \sqrt{\frac{1}{D_1} + \frac{1}{D_0}} = \sqrt{\frac{1}{84} + \frac{1}{71}} = 0.1612$$

$$\therefore \quad z_{cal} = \frac{0.3784}{0.1612} = 2.3474$$

$$\because \quad \alpha = 0.05, \quad z_{tab} = 1.96$$

Conclusion?

$$z_{cal} > z_{tab} \quad \therefore$$

Reject H_0; the baseline CS antibody status is associated with Malaria incidence.

1b):

The number of malaria cases and the person time at risk was tabulated for the malaria vaccine arm and the control arm, in each CS antibody stratum (as shown in the following page). Calculate the crude RR comparing the vaccine arm to the control arm, and the RR for each strata of baseline CS antibody status.

Baseline anti-CS antibodies concentration	Number of events (Malaria vaccine)	Months at risk (Malaria vaccine)	Number of events (Control)	Months at risk (control)
< median	48	161.1	36	111.1
≥ median	30	171.8	41	164.0

Based on this table and the previous table:

i) Do you think it might be important to adjust for baseline CS antibody status when comparing malaria incidence between the malaria vaccine and control groups? Why or why not?

ii) Do you expect the adjusted rate ratio to be the same, larger or smaller than the unadjusted rate ratio?

iii) Calculate the adjusted RR

Solution for 1b):

We can first assign variables to make it easier to identify the numbers that we are working with:

Baseline anti-CS antibodies concentration	Number of events (Malaria vaccine)	Months at risk (Malaria vaccine)	Number of events (Control)	Months at risk (control)
< median	48=d11	161.1=T11	36=d01	111.1=T01
≥ median	30=d10	171.8=T10	41=d00	164.0=T00

The Crude Rate Ratio comparing the Malaria vaccine group to the control group is:

$$CrudeRR = \frac{(d11 + d10) * (T01 + T00)}{(d01 + d00) * (T11 + T10)}$$

$$\therefore \quad CrudeRR = \frac{(48 + 30) * (111.1 + 164.0)}{(36 + 41) * (161.1 + 171.8)} = \frac{78 * 275.1}{77 * 332.9} = \frac{21457.8}{25633.3} = 0.8371$$

The Crude Rate ratio for the Malaria vaccine group is as follows:

$$RR_1 = \frac{d11 * T01}{d01 * T11} = \frac{48 * 111.1}{36 * 161.1} = \frac{5332.8}{5799.6} = 0.9195$$

On the other hand, the Crude Rate ratio for the control group is as follows:

$$RR_0 = \frac{d10 * T00}{d00 * T10} = \frac{30 * 164.0}{41 * 171.8} = \frac{4920.0}{7043.8} = 0.6985$$

Adding the results to the original table gives us the following:

Baseline anti-CS antibodies concentration	Number of events (Malaria vaccine)	Months at risk (Malaria vaccine)	Number of events (Control)	Months at risk (control)	Rate Ratio
$<$ median	48	161.1	36	111.1	$RR_1 = 0.9195$
$\geq$ median	30	171.8	41	164.0	$RR_0 = 0.6985$

Therefore the answer to 1b)i) is this: *It is important to adjust for baseline CS antibody status when comparing Malaria incidence between the vaccine and control groups. This is because there is significant deviation of each stratum's crude rate ratio from the overall crude rate ratio; they also differ significantly from each other.*

Additionally, based on the results above, the answer to 1b)ii) is this: *The adjusted Rate Ratio is expected to be less because we are se[parating and reducing the effect of the '$<$ median' stratum from the estimated effect of exposure to a Malaria vaccine.*

The adjusted RR was calculated as follows:

- The weights for each stratum were obtained

  The weight for the '$<$ *median*' stratum was:

$$W_1 = \frac{d01 * T11}{T11 + T01} = \frac{36 * 161.1}{161.1 + 111.1} = \frac{5799.6}{272.2} = 21.3064$$

  and the weight for the '$\geq$ *median*' stratum was:

$$W_0 = \frac{d00 * T10}{T10 + T00} = \frac{41 * 171.8}{171.8 + 164.0} = \frac{7043.8}{335.8} = 20.9762$$

- Q was obtained for each stratum

  Q for the '$<$ *median*' stratum was:

$$Q_1 = W_1 * RR_1 = 21.3064 * 0.9195 = 19.5912$$

  and Q for the '$\geq$ *median*' stratum was:

$$Q_0 = W_1 * RR_1 = 20.9764 * 0.6985 = 14.6520$$

- The sum of the weights was obtained

$$R = \sum_{i=1}^{m} W_i = W_1 + W_0 = 21.3064 + 20.9764 = 42.2828$$

- The adjusted RR was calculated

$$RR_{adj} = RR_{MH} = \frac{\sum_{i=1}^{m} Q_i}{R} = \frac{Q_1 + Q_0}{R} = \frac{19.5912 + 14.6520}{42.2828} = \frac{34.2432}{42.2828} = 0.8099$$

Filling a table with our results we get the following:

	Baseline anti-CS antibodies concentration	Weights (W_i)	$Q_i(W_i RR_i)$	Rate Ratio
	$<$ median	$W_1 = 21.3064$	$Q_1 = 19.5912$	$RR_1 = 0.9195$
	$\geq$ median	$W_0 = 20.9764$	$Q_0 = 14.6520$	$RR_0 = 0.6985$
TOTALS:		$R = 42.2828$	$\sum_{i=1}^{m} Q_i = 34.2432$	$RR_{adj} = 0.8099$

QUESTION 2:

The Whitehall cohort study was set up to examine risk factors for mortality in male government employees (civil servants) working around Whitehall, London. Employees were recruited between 1967 and 1970. Information on exposure to selected risk factors was obtained by a self-administered questionnaire and a screening examination during this period. All participants were followed at the National Health Service Central Registry to identify mortality and emigration. Information on death (date and cause) was provided for those who died.

Use the dataset whitehal_new.dta. We are going to explore the effect of job grade on all cause mortality. Job grade is coded as 0= administrator/professional work ('high' job grade) , 1=clerical/other work ('low' job grade). Consider level 0 as non exposed while level 1 as the exposed group as far as grade is concerned.

Import the ".dta" file into R and export it to a folder in your laptop as a ".csv" file.

In Microsoft Excel, calculate the follow-up time for each person by subtracting their entry date from their exit date. (note we are dividing by 365.25 so that we will have follow-up time in years, rather than days)

gen pyrs = (timeout-timein)/365.25

Use the pivot table in Microsoft Excel to complete the tables below:

Table 1:

Crude Population	Events exposed	Person-years exposed	Events non-exposed	Person-years non-exposed
All				

Table 2:

Smoking status	Population	Events exposed	Person-years exposed	Events non-exposed	Person-years non-exposed
1-14 cig/day	310				
15-24 cig/day	279				
25+ cig/day	125				
Ex-smoker	646				
Never smoked	317				

a) Use R commands to obtain the CRUDE rate ratio for grade on all cause mortality.
b) Now look at the RR in each strata of smoking. You can do this one-by-one with the commands similar to that used in part (a) above.
c) Obtain the Mantel-Haenszel summary measure of grade on all cause mortality adjusted for smoking.
d) Can we combine the strata? Why do you come to this conclusion?

Solution for 2a):

Based on the pivot tables generated using Excel, data was transferred to R as follows:

```
Events.exposed <- c(182)

Person.years.exposed <- c(7265.1)

Events.non.exposed <- c(221)

Person.years.non.exposed <- c(20338.2)

Population <- c(Events.exposed + Events.non.exposed)

a <- Events.exposed * Person.years.non.exposed

b <- Events.non.exposed * Person.years.exposed

Crude_all <- c(a/b)

Table1 <- cbind(Crude_all, Population, Events.exposed)

Table1 <- cbind(Table1, Person.years.exposed, Events.non.exposed, Person.years.non.exposed)

Table1 <- as.data.frame(Table1); Table1
```

Crude_all	Population	Events.exposed	Person.years.exposed	Events.non.exposed	Person.years.non.exposed
2.30542	403	182	7265.1	221	20338.2

```
smoking.status <- c('1-14 cig/day', '15-24 cig/day', '25+ cig/day', 'Ex-smoker', 'Never smoked')

population <- c(310, 279, 125, 646, 317)

events.exposed <- c(46, 57, 21, 46, 12)

person.years.exposed <- c(1853.4, 1559.3, 556.3, 2307.9, 988.1)

events.non.exposed <- c(43, 41, 27, 89, 21)

person.years.non.exposed <- c(3161.1, 2731.4, 1380.4, 8461.8, 4603.5)

Table2 = cbind(smoking.status, population, events.exposed)

Table2 <- cbind(Table2, person.years.exposed, events.non.exposed, person.years.non.exposed)

Table2 <- as.data.frame(Table2); Table2
```

smoking.status	population	events.exposed	person.years.exposed	events.non.exposed	person.years.non.exposed
1-14 cig/day	310	46	1853.4	43	3161.1
15-24 cig/day	279	57	1559.3	41	2731.4

smoking.status	population	events.exposed	person.years.exposed	events.non.exposed	person.years.non.exposed
25+ cig/day	125	21	556.3	27	1380.4
Ex-smoker	646	46	2307.9	89	8461.8
Never smoked	317	12	988.1	21	4603.5

Solution for 2b):

A breakdown of the respective rate ratio for each stratum was obtained as follows:

```
d <- events.exposed * person.years.non.exposed

e <- events.non.exposed * person.years.exposed

RR <- c(d/e)

StrataResults <- cbind(smoking.status, RR)

StrataResults <- as.data.frame(StrataResults); StrataResults
```

smoking.status	RR
1-14 cig/day	1.82456127142825
15-24 cig/day	2.43526723216953
25+ cig/day	1.92997383506102
Ex-smoker	1.89501911120134
Never smoked	2.66225223010973

Solution for 2c):

The Ratio was calculated as follows:

```
f <- events.non.exposed * person.years.exposed

g <- person.years.non.exposed + person.years.exposed

Weights <- c(f/g)

Q <- c(sum(Weights * RR))

StrataResults <- cbind(StrataResults, Weights, Q)

StrataResults <- as.data.frame(StrataResults); StrataResults
```

smoking.status	RR	Weights	Q
1-14 cig/day	1.82456127142825	15.893150	126.2732
15-24 cig/day	2.43526723216953	14.899970	126.2732
25+ cig/day	1.92997383506102	7.755512	126.2732
Ex-smoker	1.89501911120134	19.072314	126.2732
Never smoked	2.66225223010973	3.710941	126.2732

```r
R <- c(sum(Weights))

MH.adjRR <- c(Q/R)

AdjResults <- cbind(Crude_all, Q, R, MH.adjRR)

AdjResults <- as.data.frame(AdjResults); AdjResults
```

Crude_all	Q	R	MH.adjRR
2.30542	126.2732	61.33189	2.058851

Solution for 2d):

Looking at the data above, we can conclude that we should not combine the strata because:

1. The MH adjusted rate ratio differs from the crude by 0.25
2. The rate ratio of each stratum is quite different from the others; furthermore, the stratum rate ratio differs from both the crude rate ratio and the adjusted rate ratio