MODERN EPIDEMIOLOGY

ASSIGNMENT 3

Hephaes Chuen Chau (M116 I54105029)

2022-03-16

Contents

Question 1 **1**
(a) . . . 1
(b) . . . 2
(c) . . . 3
(d) . . . 3

Question 2 **3**
(a) . . . 3
(b) . . . 3
(c) . . . 3
(d) . . . 3

Question 3 **4**
(a) . . . 4
(b) . . . 4
(c) . . . 4

Question 1

(a)

We have the data set as follows:

```
row1 <- c("0-0.99", 3642, 390)
row2 <- c("1-4.99", 1504, 181)
row3 <- c("5+", 1320, 222)
data <- as.data.frame(rbind(row1,row2,row3))
colnames(data) <- c("Radiation Dose (rem)","Size of Cohort at Baseline", "New Cancer Cases")
data$`Size of Cohort at Baseline` <- as.numeric(data$`Size of Cohort at Baseline`)
data$`New Cancer Cases` <- as.numeric(data$`New Cancer Cases`)
```

Cumulative Cancer Incidence is calculated by dividing the total size by the number of new cases, as follows:

```
data$Cumulative_Cancer_Incidence <- data$`New Cancer Cases`/data$`Size of Cohort at Baseline`
data
```

```
##      Radiation Dose (rem) Size of Cohort at Baseline New Cancer Cases
## row1               0-0.99                       3642              390
## row2               1-4.99                       1504              181
## row3                   5+                       1320              222
```

```
##      Cumulative_Cancer_Incidence
## row1                   0.1070840
## row2                   0.1203457
## row3                   0.1681818
```

Relative risk (of a group relative to lowest radiation group) is calculated by dividing the cumulative cancer incidence of the group by the cumulative cancer incidence of the lowest radiation group, as follows:

```
data$`Relative Risk` <- data$Cumulative_Cancer_Incidence/0.1070840
data$`Odds Ratio (Comparing Cases to Noncases)` <-
  (data$`New Cancer Cases`*(3642-390))/
  (390*(data$`Size of Cohort at Baseline`-data$`New Cancer Cases`))
data
```

```
##      Radiation Dose (rem) Size of Cohort at Baseline New Cancer Cases
## row1               0-0.99                       3642              390
## row2               1-4.99                       1504              181
## row3                   5+                       1320              222
##      Cumulative_Cancer_Incidence Relative Risk
## row1                   0.1070840      1.000000
## row2                   0.1203457      1.123844
## row3                   0.1681818      1.570560
##      Odds Ratio (Comparing Cases to Noncases)
## row1                                 1.000000
## row2                                 1.140787
## row3                                 1.685918
```

To calculate the last column:

```
data$`Odds Ratio (Comparing Cases to Population)` <-
  (data$`New Cancer Cases`*3642)/(data$`Size of Cohort at Baseline`*390)
data
```

```
##      Radiation Dose (rem) Size of Cohort at Baseline New Cancer Cases
## row1               0-0.99                       3642              390
## row2               1-4.99                       1504              181
## row3                   5+                       1320              222
##      Cumulative_Cancer_Incidence Relative Risk
## row1                   0.1070840      1.000000
## row2                   0.1203457      1.123844
## row3                   0.1681818      1.570560
##      Odds Ratio (Comparing Cases to Noncases)
## row1                                 1.000000
## row2                                 1.140787
## row3                                 1.685918
##      Odds Ratio (Comparing Cases to Population)
## row1                                   1.000000
## row2                                   1.123844
## row3                                   1.570559
```

(b)

It can be seen that the odds ratio using noncases deviates to a larger extent than the odds ratio using population as control from the relative risk. That is because we know that when using the total population as the control group, odds ratio calculated is the same as the relative risk, due to the following formula (equation 3.14 in the textbook)

$$\text{OR}_{\text{exp}} = \frac{\text{Odds}_{\text{ exp cases}}}{\text{Odds}_{\text{ exp total population}}} = \frac{\left(\frac{a}{c}\right)}{\left(\frac{a+b}{c+d}\right)} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)} = \text{RR}$$

for a contingency table in the following form:

Exposure	Cases	Noncases	Total Population (Cases + Noncases)
Present	a	b	a + b
Absent	c	d	c + d

(c)

Only statement 4 is correct. That is because OR and RR are related to each other through (equation 3.3 from the textbook):

$$\begin{aligned}\text{OR} = \frac{\left(\frac{q_+}{1-q_+}\right)}{\left(\frac{q_-}{1-q_-}\right)} &= \frac{q_+}{1-q_+} \times \frac{1-q_-}{q_-} \\ &= \frac{q_+}{q_-} \times \left(\frac{1-q_-}{1-q_+}\right)\end{aligned}$$

where q_+ is the incidence in the exposed and q_- the incidence in the unexposed. Since now $RR > 1$, this implies that $q_+ > q_-$ and therefore the bias term in the above equation is larger than 1. This means OR deviates from 1 to a larger extent than RR.

(d)

Dose-response relationship is demonstrated.

Question 2

(a)

Case-cohort study

(b)

Relative risk is obtained. That is because we are not using non-cases as control but a subcohort as control in a case-cohort study. This situation is similar to what we have done in calculating the last column in question 1(a).

(c)

Yes. That's because case-cohort study allows for the relationship between exposure and multiple outcomes to be probed (as discussed in assignment 1).

(d)

We use the equation

$$\%\,\text{Pop AR} = \frac{p_e \times (\text{RR} - 1)}{p_e \times (\text{RR} - 1) + 1} \times 10$$

Now, p_e is the population prevalence of the exposure of interest. The exposure of interest here is having such high IL-6 values as to belong to the highest quintile. This exposure by definition can only have 20%

prevalence in any population. Therefore, we compute the required percentage population attributable risk as follows:

```
percentage_PAR <- (0.2*(1.9-1))/(0.2*(1.9-1)+1) * 100
sprintf("The required percentage AR is: %f percent", percentage_PAR)
```

```
## [1] "The required percentage AR is: 15.254237 percent"
```

Question 3

(a)

Conventional case-based case-control study

(b)

By definition, we have

```
ORless55 <- (18/(163-18))/(6/(231-6))
ORmore55 <- (52/(237-52))/(16/(165-16))
sprintf("The OR for age group with age less than or equal to 55 is %f",ORless55)
```

```
## [1] "The OR for age group with age less than or equal to 55 is 4.655172"
```

```
sprintf("The OR for age group with age larger than 55 is %f",ORmore55)
```

```
## [1] "The OR for age group with age larger than 55 is 2.617568"
```

(c)

The odds of having HCV in the past in the B-NHL group is higher than odds of having HCV in the past in the patients in the same hospital unaffected by B-NHL by 2.618 times.