**Individual Data Analysis Exam Part 1**

EPID600: Principles of Epidemiology for Public Health

**1.**

**Table 1**.Baseline Characteristics and Prevalence of Cancer

|  |  |
| --- | --- |
| **Variable** | **Value** |
| Sex - no. (%) |  |
| Male | 479 (47.9) |
| Female | 521 (52.1) |
| Age - yr |  |
| Mean | 47.15 ± 13.22 |
| Range | 19-77 |
| Age group - no. (%) |  |
| <45 | 379 (37.9) |
| ≥45 | 621 (62.1) |
| Race - no. (%) |  |
| White | 680 (68) |
| Non White | 320 (32) |
| Poverty Measure - no. (%) |  |
| At or Above 200% Federal Poverty Guideline | 705 (70.5) |
| Below 200% Federal Poverty Guideline | 295 (29.5) |
| Education - no. (%) |  |
| < High School | 185 (18.5) |
| ≥ High School | 815 (81.5) |
| Follow-up Period - yr |  |
| Mean | 9.72 |
| Range | .04-10.01 |
| Lost to Follow-up - no. (%) |  |
| No | 947 (94.7) |
| Yes | 53 (5.3) |
| Virus Status (at enrollment) - no. (%) |  |
| Unexposed | 741 (74.1) |
| Exposed | 259 (25.9) |
| Cancer - no. (%) |  |
| No | 943 (94.3) |
| Yes | 57 (5.7) |

**Prevalent variables** – Virus Status, Poverty, Education, Race, Age, Sex

**Incident variables** – Cancer Diagnosis, Lost to follow-up, Follow-up period (person years)

**2.**

The cohort consisted of 1000 healthy patients without a history of cancer diagnosis from UNC hospital affiliated clinics. We assessed whether exposure to UNC25 viral infection predicted cancer diagnosis over the follow-up a period of 10.01 years (with a mean of 9.72 years). There was roughly an even split between men and women, and 62.1% of the participants were 45 or older at the start of the study. 68% of the patients were white, 81.5% had a high school education or more, 70.5% and were above the federal poverty line. 25.9% of the patients had exposure to virus and 5.7% of the patients were diagnosed with cancer. 5.3% of the patients were lost to follow-up.

**3.**

a.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2a**. Measures of Occurrence Results (**including those lost to follow up**) | | | | | | | | |
|  | **Cancer** | **No Cancer** | **Total** | **Person-years** | **Risk** | **Risk per 100 persons** | **Rate** | **Rate per 1,000 person- years** |
| **Virus** | *33* | *226* | *259* | *2512* | *.13* | *13* | *.013* | *13* |
| **No Virus** | *24* | *717* | *741* | *7207* | *.03* | *3* | *.003* | *3* |
| **Total** | *57* | *943* | *1000* | *9272* | *.06* | *6* | *.057* | *57* |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2b**. Measures of Occurrence Results (**excluding those lost to follow up**) | | | | | | | | |
|  | **Cancer** | **No Cancer** | **Total** | **Person-years** | **Risk** | **Risk per 100 persons** | **Rate** | **Rate per 1,000 person- years** | |
| **Virus** | *25* | *216* | *241* | *2350* | *.10* | *10* | *.011* | *11* | |
| **No Virus** | *24* | *682* | *706* | *7006* | *.03* | *3* | *.003* | *3* | |
| **Total** | *49* | *898* | *947* | *9356* | *.05* | *5* | *.052* |  | |

b.

**Risk Difference** = Risk Exposed – Risk Unexposed = 10%.

The risk of cancer was 10 cases per 100 persons higher in the group exposed to the virus than in the group unexposed to the virus.

**Risk Ratio**=Risk Exposed/Risk Unexposed= 4.33

Among those exposed to the virus, the risk of being diagnosed with cancer was 4.33 times as likely compared with those not exposed to the virus.

**Rate Difference**= Rate Exposed-Rate Unexposed= 10

Among those exposed to the virus, the rate cancer diagnosis was 10 cases per 1000 person years higher compared with those not exposed to the virus.

**Rate Ratio**=Rate Exposed / Rate Unexposed =4.33

The rate of cancer among those exposed to the virus was 4.33 times the rate of cancer of those not exposed to the virus.

c.

**Risk Difference** (not including those lost to follow-up) =7%

This means that the risk of cancer was 7 cases per 100 persons higher in the group exposed to the virus than in the group unexposed to the virus.

**Risk Ratio**=Risk Exposed/Risk Unexposed= 3.33

Among those exposed to the virus, the risk of being diagnosed with cancer was 3.33 times as likely compared with those not exposed to the virus.

**Rate Difference**= Rate Exposed-Rate Unexposed= 8

Among those exposed to the virus, the rate cancer diagnosis was 8 cases per 1000 person years higher compared with those not exposed to the virus.

The **Rate Ratio** not including those lost to follow-up is 3.67. This means that the rate of cancer among those exposed to the virus was 3.67 times the rate of cancer of those not exposed to the virus.

The risk ratio, risk difference, rate ratio and rate difference were lower not including those lost to follow-up. 8 out of 53 patients lost to follow-up were diagnosed with cancer. Considering only 57 out 1000 patients were diagnosed with cancer removing these individuals clearly has an impact on the risk and rate calculations.

Appendix

dat = read\_excel("IDA\_Dataset+8.7.18.xlsx")  
dat=dat%>%select(-c(X\_\_1,X\_\_2))  
dat2=dat%>%select(c(CANCER,VIRUS,PERSONDAYS,LTF))  
dat2=dat2%>%mutate(personyrs=PERSONDAYS/365)

taba=table(dat$VIRUS,dat$CANCER)[2:1,2:1]  
x=c("Exposed","Unexposed")  
y=c("Cancer","No Cancer")  
row.names(taba)=x  
colnames(taba)=y  
kable(taba)

|  |  |  |
| --- | --- | --- |
|  | Cancer | No Cancer |
| Exposed | 33 | 226 |
| Unexposed | 24 | 717 |

dat2%>%filter(VIRUS==1)%>%summarise(pyrs=sum(personyrs))

## # A tibble: 1 x 1  
## pyrs  
## <dbl>  
## 1 2512.

dat2%>%filter(VIRUS==0)%>%summarise(pyrs=sum(personyrs))

## # A tibble: 1 x 1  
## pyrs  
## <dbl>  
## 1 7207.

dat3=dat2%>%filter(LTF==0)  
length(dat3$LTF)

## [1] 947

tabb=table(dat3$VIRUS,dat3$CANCER)[2:1,2:1]  
row.names(tabb)=x  
colnames(tabb)=y  
kable(tabb)

|  |  |  |
| --- | --- | --- |
|  | Cancer | No Cancer |
| Exposed | 25 | 216 |
| Unexposed | 24 | 682 |

dat3%>%filter(VIRUS==1)%>%summarise(pyrs=sum(personyrs))

## # A tibble: 1 x 1  
## pyrs  
## <dbl>  
## 1 2350.

dat3%>%filter(VIRUS==0)%>%summarise(pyrs=sum(personyrs))

## # A tibble: 1 x 1  
## pyrs  
## <dbl>  
## 1 7006.