**BS852: Midterm Exam**

**Jiayuan Shi**

*Question 1.*

1. What type of study is this?

**Solution:** Case control.

2. What is problematic with this analysis?

**Solution:** This study has matched cases and controls on confounders but analyze the data ignoring matching will result in a biased estimate of association. It is possible that this estimate will be biased toward to null and so we may expect that the true effect will be stronger than this. Therefore, unmatched analysis is not a valid analytical approach to study with matching.

3. Complete the table below in which the data are summarized by matched pairs.

**Solution:**

|  |  |  |
| --- | --- | --- |
|  | Control | |
| Case | Tonsillectomy | None |
| Tonsillectomy | 37 | 7 |
| None | 15 | 26 |

4. The OR for the association turns out to be OR =0.467, with a Chi-square test 2.901 on 1 degrees of freedom.

Compute a 95% confidence interval for the OR. Use this result to discuss the conclusion of the study based on this analysis.

**Solution:** The 95% confidence interval for the OR is,

H0: There is no association between tonsillectomy and Hodgkin’s lymphoma after matching on potential confounders.

H1: There is an association between tonsillectomy and Hodgkin’s lymphoma after matching on potential confounders.

Since the chi-square statistic is 2.901 and its p-value is 0.089. At significance level of 0.05, we cannot reject the null hypothesis and conclude that there is not a significant association between tonsillectomy and Hodgkin’s lymphoma after matching on potential confounders. With 95% confidence, we estimate that the true OR lies between 0.194 and 1.122. This interval contains the null value of 1, and thus agreeing with our final conclusion on non-significant association between tonsillectomy and Hodgkin’s lymphoma after matching on potential confounders.

People who had Tonsillectomy have 6.6 times the odds of Hodgkin’s Lymphoma compared to those did not have Tonsillectomy after matching on potential confounders.

5. Complete the 2 tables of matched pairs and explain how the missing numbers were computed. Use the completed tables to test the hypothesis that there is an interaction between sex and tonsillectomy on the chance for Hodgkin’s disease.

**Solution:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Control (male) | | N. pairs |  | Control (Female) | | N. pairs |
| Case (male) | Tonsillectomy | None |  | Case (female) | Tonsillectomy | None |  |
| Tonsillectomy | 28 | 2 | 30 | Tonsillectomy | 9 | 5 | 14 |
| None | 2 | 13 | 15 | None | 13 | 13 | 26 |
| N. pairs | 30 | 15 | 45 | N. pairs | 22 | 18 | 40 |

For male with both Hodgkin’s disease and no Hodgkin’s disease,

* who had tonsillectomy were **28** pairs: 30-2=28
* who did not have tonsillectomy were **13** pairs: 26-13=13 (toal pairs in this situation minus the given female’s pairs)

For male both with Hodgkin’s disease who had tonsillectomy and without Hodgkin’s disease who did not have tonsillectomy were **2** pairs (given).

For male both with Hodgkin’s disease who did not have tonsillectomy and without Hodgkin’s disease who had tonsillectomy were **2** pairs:

1. # of pairs with Hodgkin’s disease who did not have tonsillectomy: 45-30=15 (45 pairs male siblings minus 30 pairs with Hodgkin’s disease who had tonsillectomy)
2. 15-13=2 (# of male pairs with Hodgkin’s disease who did not have tonsillectomy minus pairs with both Hodgkin’s disease and no Hodgkin’s disease who did not have tonsillectomy)

For female with both Hodgkin’s disease and no Hodgkin’s disease,

* who had tonsillectomy were **9** pairs: 37-28=9 (toal pairs in this situation minus the given male’s pairs)
* who did not have tonsillectomy were **13** pairs (given)

For female both with Hodgkin’s disease who had tonsillectomy and without Hodgkin’s disease who did not have tonsillectomy were **5** pairs: 7-2=5 (toal pairs in this situation minus the given male’s pairs)

For female both with Hodgkin’s disease who did not have tonsillectomy and without Hodgkin’s disease who had tonsillectomy were **13** pairs: 15-2=13 (toal pairs in this situation minus the given male’s pairs).

H0: There is no interaction between sex and tonsillectomy. (ORs are the same between the sex groups)  
H1: There is interaction between sex and tonsillectomy. (ORs are not the same between the sex groups)

|  |  |  |
| --- | --- | --- |
| Discordant pairs | Male | Female |
| Yes-No | 2 | 5 |
| No-Yes | 2 | 13 |

We want to see if the proportions of discordant pairs are the same across age groups, we can use Pearson Chi-square test.

Pearson Chi-square test statistic = , (degree of freedom=1)

The chi-square statistic is 0.745 and its p-value is 0.388. At significance level of 0.05, we do not have evidence to reject the null and thus we conclude that there is no interaction between sex and tonsillectomy. It is likely that we do not have enough discordant pairs to detect possible interactions since two cells have counts less than 5.

6. Can you describe possible limitations of the analysis in part 5?

**Solution:** We cannot compare if male have more Hodgkin’s disease than women because we are controlling sex. Further, when subjects are matched, comparison of tonsillectomy is only appropriately done withinmatched sets.

7. Could the investigators use a similar approach to test whether there is an interaction between age and tonsillectomy and whether there is an interaction between race and tonsillectomy?

**Solution:** Yes.

8. Provide overall conclusions of the analysis.

**Solution:** The type of study is case control, and it has matched cases and controls on confounders so we use matched analysis.

After matching on potential confounders, there is not a significant association between tonsillectomy and Hodgkin’s lymphoma (chi-square=2.90, p-value=0.089). With 95% confidence, we estimate that the true OR lies between 0.194 and 1.122. This interval contains the null value of 1, and thus agreeing with our final conclusion on non-significant association between tonsillectomy and Hodgkin’s lymphoma.

People who had Tonsillectomy have 6.6 times the odds of Hodgkin’s Lymphoma compared to those did not have Tonsillectomy after matching on potential confounders.

When matching pairs stratified by sex, we conclude that there is no interaction between sex and tonsillectomy (chi-square=0.745, p-value=0.388). It is likely that we do not have enough discordant pairs to detect possible interactions since two cells have counts less than 5.

*Question 2*

1. What type of study is this?

**Solution:** Case control.

2. Test the null hypothesis that there is no interaction between sex and exposure to regular alcohol use, using Breslow-Day test of interaction.

**Solution:**

H0: There is no interaction between sex and exposure to regular alcohol use (i.e. all OR’s  
are the same).  
H1: There is interaction between sex and exposure to regular alcohol use (i.e. at least 1 of the two OR’s are not the same).

OR for males = 3.92, and OR for females = 1.79.

Breslow-Day Test:  
The chi-square statistic (degree of freedom = 1) is 18.17 and its p-value is 2.02e-05 < 0.05 . At significance level of 0.05, we have significant evidence to reject the null hypothesis, and therefore, we can conclude that there is significant interaction between sex and exposure to regular alcohol use.

3. Write down stratum-specific odds ratios, and 95% confidence intervals for each sex-stratum. Do the results show that the sex-specific associations between exposure to alcohol use and risk of falling are significant? Describe in words the effects of sex on the association between regular alcohol consumption and risk of falling.

**Solution:**

$OR.strata

est lower upper

1 3.919551 2.976795 5.203186

2 1.792038 1.396387 2.298184

$chisq.strata

test.statistic df p.value

1 108.59287 1 0.00000e+00

2 22.70972 1 1.88412e-06

So for males, chi-square=108.59, p-value<0.0001. OR is 3.92 with 95% CI (2.98, 5.20).

For females, chi-square=22.71, p-value<0.0001. OR is 1.79 with 95% CI (1.40, 2.30).

The results show that the sex-specific associations between exposure to alcohol use and risk of falling are significant. We analyze by splitting it into males and females in the followings,

H0: There is association between exposure to alcohol use and risk of falling are significant among males, OR=1  
H1: There is an association between exposure to alcohol use and risk of falling are significant among males, OR ≠ 1

The above chi-square statistic for males is 108.59 with p-value is < 0.05. At significance level of 0.05, we have observed evidence to reject the null hypothesis, and therefore, we can conclude that there is a significant association between exposure to alcohol use and risk of falling for males. With 95% confidence, we estimate that the true OR for males lies between 2.98 and 5.20. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling for males.

Males with age 65 and older who were exposure to alcohol use had 3.92 times the odds of falling compared with males who were not exposure to alcohol use.

H0: There is association between exposure to alcohol use and risk of falling are significant among females, OR=1  
H1: There is an association between exposure to alcohol use and risk of falling are significant among females, OR ≠ 1

The above chi-square statistic for females is 22.71 with p-value is < 0.05. At significance level of 0.05, we have observed evidence to reject the null hypothesis, and therefore, we can conclude that there is a significant association between exposure to alcohol use and risk of falling for females. With 95% confidence, we estimate that the true OR for females lies between 1.40 and 2.30. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling for females.

Females with age 65 and older who were exposure to alcohol use had 1.79 times the odds of falling compared with females who were not exposure to alcohol use.

4. Write down the sex adjusted odds ratio, the 95% confidence interval, and the Mantel-Haenszel chi square and use the results to show whether there is a significant association between exposure to alcohol and risk of falling, after adjusting for sex.

**Solution:**

$OR.mh

est se lower upper

1 2.612077 0.08990484 2.190074 3.115395

$chisq.mh

test.statistic df p.value

1 116.4953 1 3.702479e-27

Mantel-Haenszel chi-square=116.50, p-value=3.70e-27.

Adjusted OR is 2.61 with 95% CI (2.19, 3.12).

H0: There is no association between exposure to alcohol and risk of falling, after adjusting for sex, OR=1  
H1: There is an association between exposure to alcohol and risk of falling, after adjusting for sex, OR ≠ 1

The estimated adjusted OR is 2.61. The Mantel Haenszel chi-square statistic is 116.50 (degree of freedom = 1), and its p-value is 3.70e-27 which is smaller than 0.05. At significance level of 0.05, we have sufficient observed evidence to reject the null hypothesis and conclude that there is indeed an association exposure to alcohol and risk of falling, after adjusting for sex. The 95% confidence interval is (2.19, 3.12). With 95% confidence, we estimate that the true OR for falling lies between 2.19 and 3.12. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling, after adjusting for sex.

After adjusting for sex, people with age 65 and older who were exposure to alcohol use had 2.61 times the odds of falling compared with people who were not exposure to alcohol use.

5. What are the overall conclusions based on these analyses?

**Solution:** The type of study is case control.

From the Breslow-Day test, at significance level of 0.05, we can conclude that there is significant interaction between sex and exposure to regular alcohol use (chi-square=18.17, p-value is 2.02e-05).

The results show that the sex-specific associations between exposure to alcohol use and risk of falling are significant. We analyze by splitting it into males and females in the followings,

* For males, there is a significant association between exposure to alcohol use and risk of falling (chi-square=108.59, p-value is < 0.05). With 95% confidence, we estimate that the true OR for males lies between 2.98 and 5.20. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling for males. Males with age 65 and older who were exposure to alcohol use had 3.92 times the odds of falling compared with males who were not exposure to alcohol use.
* For females, there is a significant association between exposure to alcohol use and risk of falling (chi-square=22.71, p-value is < 0.05). With 95% confidence, we estimate that the true OR for females lies between 1.40 and 2.30. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling for females. Females with age 65 and older who were exposure to alcohol use had 1.79 times the odds of falling compared with females who were not exposure to alcohol use.

Since there is an interaction between sex and exposure to regular alcohol use, we should use the stratum-specific analysis above. However, if we adjust for sex, we have sufficient observed evidence to reject the null hypothesis and conclude that there is indeed an association exposure to alcohol and risk of falling (Mantel Haenszel chi-square=116.50, p-value=3.70e-27). The 95% confidence interval for OR=2.61 is (2.19, 3.12). With 95% confidence, we estimate that the true OR for falling lies between 2.19 and 3.12. This interval excludes the null value of 1, and thus agrees the significant association between exposure to alcohol use and risk of falling, after adjusting for sex.

*Question 3*

1. Write down the formulas for each of the three models

**Solution:**

Model1:

Model2:

Model3:

2. Does age confound the relationship (a) between MDR and “male”, (b) between MDR and “homeless”? Explain your answers.

**Solution:**

(a)

The two measures differ by less than 10%, so we conclude that age does not confound the relationship between MDR and “male”.

(b)

The two measures differ by less than 10%, so we conclude that age does not confound the relationship between MDR and “homeless”.

3. Which model do you think the investigators should choose as their final model? Explain your answer, showing any workings you did.

**Solution:** I think the investigators should choose Model1 as their final model, because with 95% confidence, all of the age (p= 4.03e-05), male (p= 0.0256) and homeless (p= 0.0077) have significant association with MDR. Therefore, we should include all the variables in the model.

In the Model1,

OR. Age\_lin = exp(-0.014362) = 0.986

OR. Male= exp(0.242044) = 1.274

OR. Homeless = exp(0.456893) = 1.579

For each year of age, the odds of MDR-TB increases by 0.986 times, controlling for male and homeless. Males have 1.274 times the odds of MDR-TB as females, controlling for age and homeless. Homeless have 1.579 times the odds of MDR-TB as non-homeless, controlling for age and male.

4. Using model 1, report the factors that are significantly associated with having MDR-TB. List the odds ratios and confidence intervals and describe these associations in words

**Solution:**

H0: all the variables are not associated with having MDR-TB; H1: at least one of the variables is associated with having MDR-TB.

The model chi-square = 3022.6- 2996.0 = 26.6, with 3 degrees of freedom, p-value < 0.01. Therefore, at least one of the variables in the model is significantly associated with having MDR-TB, and we do the tests below.

Age:

OR. Age\_lin = exp(-0.014362) = 0.986

H0: age is not associated with having MDR-TB; H1: age is associated with having MDR-TB.

Chi-square = (Z values)^2 = (-4.106)^2 = 16.8592

CI for OR:

There is a significant association between age and MDR-TB (chi-square=16.8592, p-value= 4.03e-05<0.05). With 95% confidence, we estimate that the true OR for males lies between 0.979 and 0.993. This interval excludes the null value of 1, and thus agrees the significant association. For each year of age, the odds of MDR-TB increases by 0.986 times, controlling for male and homeless.

Male:

OR. Male= exp(0.242044) = 1.274

H0: male is not associated with having MDR-TB; H1: male is associated with having MDR-TB.

Chi-square = 2.232 ^2 = 4.9818

CI for OR:

There is a significant association between male and MDR-TB (chi-square=4.9818, p-value = 0.0256<0.05). With 95% confidence, we estimate that the true OR for males lies between 1.030 and 1.576. This interval excludes the null value of 1, and thus agrees the significant association. Males have 1.274 times the odds of MDR-TB as females, controlling for age and homeless.

Homeless:

OR. Homeless = exp(0.456893) = 1.579

H0: homeless is not associated with having MDR-TB; H1: homeless is associated with having MDR-TB.

Chi-square = 2.665^2 = 7.1022

CI for OR:

There is a significant association between homeless people and MDR-TB (chi-square=7.1022, p-values= 0.0077 <0.05). With 95% confidence, we estimate that the true OR for males lies between 1.128 and 2.209. This interval excludes the null value of 1, and thus agrees the significant association. Homeless have 1.579 times the odds of MDR-TB as non-homeless, controlling for age and male.

Therefore, age (chi-square=16.8592, p-value <0.05), male (chi-square=4.9818, p-value <0.05) and homeless (chi-square=7.1022, p-values <0.05) all associated with having MDR-TB significantly.

5. Is this a better model than model 1? Show any workings.

**Solution:**

H0: education level is not associated with having MDR-TB; H1: education level is associated with having MDR-TB.

LR test:

Chi-square = -2lnR = 2996.0-2992.1 = 3.9 with 4 df, p-value = 0.4197.

Since the p-value from both tests is > 0.05, we can reject the H0, and conclude that education level is not associated with having MDR-TB. Also, for each of the education levels (education2, education3, education4 and education5) in the output, we can see their p-values are larger than 0.05 (p –value for education2: 0.3128, education3: 0.5887, education4: 0.2753 and education5: 0.4911), so this also proves that the education levels are not associated with having MDR-TB. Therefore, this model is not a better model than model 1.