**Confounding and effect modification practical**

In this practical session we will learn how to use Stata to deal with **confounding and effect modification** in analyses. We will use the dataset that you have been using throughout the course so far. We will look at associations between two variables in this dataset and whether there is suggestion of confounding or effect modification by other variables.

This practical is split into three sections. For each section there is a video that shows you how to use the Stata commands. You might choose to watch the video first and then work through the practical exercise by yourself. Or you might instead choose to work through the practical alongside watching the video.

**A. Dealing with confounding in associations with a binary outcome**

**B. Dealing with effect modification in associations with a binary outcome**

**C. Dealing with multiple confounders or effect modifiers in associations with a binary outcome**

Remember!

* Open up the lecture notes/slides which match this session
* Make sure you save your commands in a do file
* Some of the Stata commands you will use in this session are ones you will have seen before, so you should open do files used in previous sessions.

**A. Dealing with confounding in associations with a binary outcome variable**

In this first section we are going to look at the association between physical activity (the exposure) and heart attack (the outcome). Physical activity is represented by variable **physact** and incident heart attack by **heart\_attack.**

***A1: Examining the variables***

We should first examine these two key variables. We will do this using the **tab** command in Stata.

. tab physact

3 categories of |

physical activity |

(1=low, 3=high) | Freq. Percent Cum.

Group 1 – low | 1,321 42.23 42.23

Group 2 – medium | 1,051 33.60 75.83

Group 3 – high | 756 24.17 100.0

Total | 3,128 100.0

*Q: How is physical activity distributed in our sample?*

*A: There are 3128 people with information on physical activity in our sample, 1321 (42.2%) reported low physical activity, 1051 (33.6%) report medium activity, and 756 (24.2%) report high physical activity.*

. tab heart\_attack

Incident heart attack | Freq. Percent Cum.

Not mentioned | 2,903 92.84 92.84

Mentioned | 224 7.16 100.00

Total | 3,127 100.00

*Q: How many people in our sample had a heart attack?*

*A: There are 3127 people with complete information on heart attack. 224 people in our sample had a heart attack (7.2%).*

***A2: Examining the association between physical activity and heart attack***

We can examine the association between physical activity and heart attack using a chi-squared test as we have two categorical variables. Remember, to do this we cross-tabulate these two variables and add the option **chi** to report the chi-squared statistics.

Remember when adding an option to a Stata command, it goes after the comma!

*Q: Is physical activity associated with heart attack?*

. tab heart\_attack physact, col chi

3 categories of physical

Activity (1=low, 3=high)

Incident heart attack | Group 1 – Group 2 – Group 3- Total

Not mentioned | 1,196 990 716 2,902

| 90.54 94.38 94.71 92.83

Mentioned | 125 59 40 224

| 9.46 5.62 5.29 7.17

Total | 1,321 1,049 756 3,126

| 100.00 100.00 100.00 100.00

Pearson chi2(2) = 18.2161 Pr = 0.000

`

*A: Yes, there appears to be an association between physical activity and having a heart attack. 9.5% of those reporting low physical activity had a heart attack compared to 5.6% of those reporting medium and 5.3% of those reporting high physical activity. The null hypothesis is that there is no association between physical activity and heart attack. The chi-squared statistic (fairly high) and association p value (very low, <0.001) suggest that there is little compatibility with being selected from a population where those with the three levels of physical activity had the same proportion of people with heart attacks. We can therefore conclude that there is an association between physical activity and heart attack in this sample.*

We can also express the association between physical activity and heart attack in terms of an odds ratio using the **mhodds** command in Stata. In the mhodds command we specify the outcome variable (heart\_attack) before the exposure variable (physact). The interpretation is slightly different to that presented in the lecture as physical activity has 3 categories (instead of 2).

*Q. Describe the association between physical activity and heart attack as an odds ratio and 95% confidence intervals.*

. mhodds heart\_attack physact

Score test for trend of odds with physact

(The Odds Ratio estimate is an approximation to the odds ratio for a one unit increase in physact)

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

0.712679 15.08 0.0001 0.600678 0.845563

*A. As the frequency of physical activity increases the likelihood of having a heart attack reduces by 28.7%. The confidence interval shows us that the true population odds ratio estimate is likely to lie somewhere between 0.60 and 0.85 95% of the time. The p value is very low 0.0001.*

***A3. Assessing the role of self-rated health in the association between physical activity and heart attack***

In this section we will assess whether self-rated health is a confounder of the association between physical activity and heart attack.

First self-rated health needs to be recoded into a new binary variable as we would like to have two categories comparing those with ‘very good’ and ‘good’ health, with people who have ‘fair’, ‘bad’ or ‘very bad’ health.

*Q. Generate a new self-rated health variable with two such categories. Ensure that the categories are labelled.*

Remember, it is good practice to create a new variable rather than makes changes to an existing variable. If you make a mistake you can ‘drop’ the new variable and easily start again.

*A. First we should check the coding of the existing variable. We can do this using the ‘codebook’ command.*

. codebook srh

srh Self-reported general health

Type: numeric (byte)

Label: hegenh

Range: [1,5] units:1

Unique values: 5 missing: 52/3,129

Tabulation: Freq. Numeric Label

722 1 …very good,

1,279 2 good,

862 3 fair,

169 4 bad,

45 5 or, very bad?

52 .

*Now we can use this information to create a new binary variable.*

. recode srh 1/2=0 3/5=1, gen(newsrh)

. label define srh 0 “VG or good” 1 “fair or bad”

. label values newsrh srh

. tab newsrh srh

RECODE of |

srh |

(Self-reported |

General health)|…very g good, fair, bad, or, very |Total

VG or good | 722 1,279 0 0 0 |2,001

Fair or bad | 0 0 862 169 45 |1,076

Total 722 1,279 862 169 45 |3,077

*By cross-tabbing the new and old variable we can check that the correct categories have been collapsed into the new binary variable.*

Remember, to be a confounder the variable should be associated with both the exposure and outcome. Also it should not lie on the causal pathway.

To meet the conditions of a confounder we will firstly assess whether the new binary self-rated health variable is associated with both physical activity and heart attack. Again, we can do this using a chi-squared test.

*Q. Run two Chi-squared tests to check whether self-rated health is associated with both the exposure (physical activity) and outcome (heart attack). What do you conclude?*

. tab heart\_attack newsrh, col chi

| RECODE of srh

Incident heart attack | (Self-reported general health)

| VG or good fair or bad | Total

Not mentioned | 1,928 927 | 2,855

| 96.35 86.31 | 92.85

Mentioned | 73 147 | 220

| 3.65 13.69 | 7.15

Total | 2,001 1,074 | 3,075

| 100.00 100.00 |100.00

Pearson chi2(1) = 106.0343 Pr = 0.000

*Here we can see that there is an association between self-rated health and heart attack. 3.7% of those with very good or good health had a heart attack, however 13.7% of those with fair or bad health had a heart attack. The Chi-squared statistic is high (106.03) with a low p value (<0.001). Therefore we can say that self-rated health and heart attack are associated. It appears that those with poorer health are more likely to have had a heart attack.*

. tab physact newsrh, col chi

| RECODE of srh

3 categories of | (Self-reported

Physical activity | general health)

(1=low, 3=high) | VG or good fair or bad |Total

Group1 – low | 693 595 |1,288

| 34.63 55.35 |41.87

Group2 – medium | 751 287 |1,038

| 37.53 26.70 |33.75

Group3 – high | 557 193 | 750

| 27.84 17.95 |24.38

Total | 2,001 1,075 |3,076

| 100.00 100.00 |100.00

Pearson chi2(2) = 124.0069 Pr=0.000

*Again there appears to be an association between self-rated health and physical activity. 27.84% of those with very good or good health reported high physical activity. 17.95% of those with fair or bad health reported high physical activity. Similarly, a greater percentage of people with fair or bad health reported low physical activity. The Chi-squared statistic is again high (124.01) and the p value is low (p<0.001) suggesting low compatibility with people with good or bad health having reporting the same levels of physical activity.*

*Therefore we can conclude that self-rated health is associated with both our exposure (physical activity) and outcome (heart attack). It is also unlikely to be on the causal pathway between these two variables and we therefore consider it to be a confounder of this association.*

The next step is to use the mhodds command to see what effect the inclusion of self-rated health has in our analysis of the association between physical activity and heart attack.

*Q. Assess the role of self-rated health in the association between physical activity and heart attack using the mhodds command. Interpret your findings.*

*A. We showed earlier that physical activity was associated with heart attack (OR=0.71, 95% CI: 0.60, 0.85) – for each increasing category of reported physical activity there was a lower odds of heart attack. Here we will assess the role of self-rated health in this association using the mhodds command.*

.mhodds heart\_attack physact, by(newsrh)

Score test for trend of odds with physact by newsrh

(The Odds Ratio estimate is an approximation to the odds ratio for a one unit increase in physact)

newsrh | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

VG or good | 0.776491 2.79 0.0947 0.57714 1.04471

Fair or poor| 0.842509 2.22 0.1367 0.67230 1.05580

Mantel-Haenszel estimate controlling for newsrh

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

0.817687 4.82 0.0281 0.683248 0.978578

Test for homogeneity of ORs (approx): chi2(1) = 0.18

Pr>chi2 = 0.6679

*Looking firstly at the stratified analyses we can see that the ORs reported for each stratum (good health and fair/poor health) are similar. The p value at the bottom of the output is consistent with this, suggesting that the stratum specific effects don’t differ. We can therefore look at the pooled estimate at the bottom of the output which gives the OR for the association between physical activity and heart attack, controlling for self-rated health. This estimate suggests that for each increasing level of physical activity there is reduced odds of heart attack after accounting for self-rated health (OR=0.82, 95% CI: 0.68, 0.98). This estimate lies closer to 1 (‘no effect’), therefore part of the association between physical activity and heart attack is confounded by self-rated health.*

***==================================================================================***

***B1 Assessing the role of sex in the association between physical activity and heart attack***

In this second practical we are going to apply the methods for testing for effect modification in associations involving a binary outcome as shown in part B of the lecture. Please watch part B of the lecture before attempting this practical exercise.

We will return to the association between physical activity and heart attack that we tested in part A. Here we will assess whether sex plays a role in this association as an effect modifier.

*Q. Using the mhodds command, assess whether sex plays a role in the association between physical activity and heart attack. What do you conclude?*

*A.*  *We use the same command as in section A:*

.mhodds heart\_attack physact, by(sex)

Score test for trend of odds with physact by sex

(The Odds Ratio estimate is an approximation to the odds ratio for a one unit increase in physact)

Sex | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

Men | 0.769543 6.28 0.0122 0.62702 0.94446

Women | 0.513824 18.07 0.0000 0.37799 0.69848

Mantel-Haenszel estimate controlling for sex

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

0.679529 19.75 0.0000 0.573074 0.805759

Test for homogeneity of ORs (approx): chi2(1) = 4.60

Pr>chi2 = 0.0320

*Focussing firstly on the bottom of the output on the test for homogeneity of ORs, we can see that there is suggestion that sex might be an effect modifier (p=0.0320). Therefore we do not use the pooled Mantel-Haenszel estimate in this scenario and instead look at the stratum-specific estimates in the top of the output. Here the associations are stratified by sex. For men, there is an association between physical activity and heart attack; for each increasing category of physical activity intensity there is a reduction in odds of heart attack (OR=0.77, 95% CI: 0.63, 0.94). For women the association works in a similar direction but is stronger – for each increasing category of physical activity intensity there is a reduction in odds of heart attack (OR=0.51, 95% CI: 0.38, 0.70). We can therefore conclude that sex is an effect modifier of the relationship between physical activity and heart attack, and that the association is stronger for women.*

***B2 Assessing the role of wealth in the association between physical activity and heart attack***

*Q. Assess whether wealth plays a role in the association between physical activity and heart attack following a similar approach to section B1. Use variable ‘wealth5’ in your analysis.*

*A. Here we are assessing whether wealth in quintiles (variable wealth5) might be a confounder or effect modifier in our association between physical activity and heart attack. We again use the mhodds command to test this.*

. mhodds heart\_attack physact, by(wealth5)

Score test for trend of odds with physact by wealth5

(The Odds Ratio estimate is an approximation to the odds ratio for a one unit increase in physact)

Wealth5 Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

1 0.875411 0.63 0.4269 0.63047 1.21552

2 0.643293 4.85 0.0277 0.43438 0.95267

3 0.709068 2.83 0.0926 0.47498 1.05853

4 0.666036 3.51 0.0612 0.43523 1.01923

5 0.729985 2.26 0.1329 0.48420 1.10053

Mantel-Haenszel estimate controlling for wealth5

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

0.733896 12.30 0.0005 0.617374 0.872411

Test for homogeneity of ORs (approx.): chi2(4) = 1.77

Pr>chi2 = 0.7780

*Focussing on the test for homogeneity first, the Chi-square statistic is low and the p value is high suggesting that we should not use the stratum-specific estimates – i.e. the relationship between physical activity and heart attack does not vary across the strata of wealth5. We can confirm this by viewing the similar ORs at the top of the output and their substantially overlapping CIs. We can therefore say that wealth5 is not an effect modifier of the association between physical activity and heart attack. The next step is to view the pooled estimate which shows the association between physical activity and heart attack controlling for wealth, OR=0.73 (95% CI: 0.62, 0.87). The estimate is very similar to that from the crude association. We would therefore not consider wealth5 to be a confounder of the association between physical activity and heart attack and therefore we would not use this variable further in our analyses.*

***Optional - assessing the role of age group in the association between diabetes and heart attack***

If you would like some further practice in using the mhodds command to assess the role of third variables as confounders and effect modifiers you might look to consider the association between diabetes and heart attack and whether age plays a role. You will need to create a new age group variable as the age variable in the dataset (variable name age) is currently in continuous form. We suggest coding age into the following categories – min-55, 56-65, 66-75, and 76+. This is also extra practice for generating and labelling new variables in Stata!

*Q. What role does age group play in the association between diabetes and heart attack?*

*A. First we generate the new agegroup variable from the existing variable called age and recode it into the suggested categories:*

. recode age min/55=0 56/65=1 66/75=2 76/max=3, gen(newage)

. label define agegp 0 “min-55” 1 “56-65” 2 “66-75” 3 “76+”

. label values newage agegp

. tab newage

RECODE of |

age (Age in |

years) | Freq. Percent Cum.

Min-55 | 1,216 38.86 38.86

56-65 | 986 31.51 70.37

66-75 | 679 21.70 92.07

76+ | 248 7.93 100.00

Total 3,129 100.00

*It is a good idea to examine the variable* ***diab*** *as we have not previously used this in our analysis.*

. codebook diab

diab Doctor diagnosed diabetes

type: numeric (byte)

label: bin

range: [0,1] units: 1

unique values: 2 missing.: 0/3,129

tabulation: Freq. Numeric Label

3,020 0 No

109 1 Yes

*There are 109 people in our sample with doctor diagnosed diabetes.*

*Next we examine the crude association between diabetes and heart attack. We can do this using the mhodds command.*

. mhodds heart\_attack diab

Maximum likelihood estimate of the odds ratio

Comparing diab==1 vs. diab==0

Odds Ratio chi2(1) P.chi2 [95% Conf. Interval]

2.324235 9.59 0.0020 1.341264 4.027595

*People with type 2 diabetes are more likely to report a heart attack than those who do have type 2 diabetes (OR=2.32, 95% CI: 1.34, 4.03).*

*To assess the role of age group we run the following command:*

. mhodds heart\_attack diab, by(newage)

Maximum likelihood estimate of the odds ratio

Comparing diab==1 vs. diab==0

by newage

--------------------------------------------------------------------------

newage | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------min-55 | 4.064630 5.68 0.0172 1.16419 14.19119

56-65 | 1.353846 0.24 0.6253 0.39949 4.58804

66-75 | 1.189820 0.15 0.7026 0.48704 2.90666

76+ | 9.000000 12.32 0.0004 2.03001 39.90125

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for newage

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.929751 5.70 0.0170 1.113978 3.342920

----------------------------------------------------------------

Test of homogeneity of ORs (approx.): chi2(3) = 7.89

Pr>chi2 = 0.0484

*Firstly looking at the test for homogeneity results we can see there might be some evidence of effect modification and heterogeneity of associations across strata. If we inspect the stratum specific results at the top of the output we can see that there is quite a lot of variation in the association between type 2 diabetes and heart attack by age group. More specifically, the association only seems to exist for the youngest and oldest age groups and in the same direction; those aged up to 55 years who had type 2 diabetes were much more likely to have heart attack compared to those without type 2 diabetes in this age group (OR=4.06, 95% CI: 1.16, 14.19). Also those aged 76 and over who had type 2 diabetes were much more likely to have a heart attack than those aged 76 and over without type 2 diabetes (OR=9.00, 95% CI: 2.03, 39.90). There doesn’t appear to be much of an association between diabetes and heart attack for those aged 56-75 years. We would therefore conclude that age is an effect modifier of the association between type 2 diabetes and heart attack and present the stratum specific results.*

***==================================================================================***

***C1 Considering multiple confounders using mhodds***

In the lecture part C we saw that we could consider up to four confounding variables and two effect modifiers in the mhodds command in Stata. Please ensure you have watched part C of the lecture before attempting this practical. In this practical we are going to consider more than one confounder and effect modifier in our analyses.

In this section we are going to test the association between smoking status and activities of daily living (ADLs). Please use the binary smoking variable **smok\_bin** with two categories – one with never/ex occasional smokers, and one with ex-regular/current smokers. Please also use the binary variable **ADL** that indicates whether participants had at least one ADL versus having none.

*Q. Explore these two variables and test the crude association between binary smoking status and having an ADL using the mhodds command. What do you conclude?*

*A. First we need to explore our data*

. tab smok\_bin

Binary smoking |

status | Freq. Percent Cum.

---------------+-----------------------------------

never/ex occ | 1,375 43.94 43.94

ex/current reg | 1,754 56.06 100.00

---------------+-----------------------------------

Total | 3,129 100.00

. tab ADL

Presence of |

1+ ADLs | Freq. Percent Cum.

------------+-----------------------------------

No ADLs | 2,508 80.54 80.54

1+ ADLs | 606 19.46 100.00

------------+-----------------------------------

Total | 3,114 100.00

*We can see that in our sample, more than half of the participants were ex regular or current regular smokers (56.1%, n=1754). The remainder (n=1375, 43.9%) had never smoked or had been smoked occasionally. Regarding ADLs, almost one fifth of the sample reported at least one difficulty with daily activities (606, 19.5%).*

*We next test the association between smoking and ADLs using mhodds:*

. mhodds ADL smok\_bin

Maximum likelihood estimate of the odds ratio

Comparing smok\_bin==1 vs. smok\_bin==0

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.346301 10.31 0.0013 1.122123 1.615267

----------------------------------------------------------------

*People in our dataset who are current or ex regular smokers are more likely to have at least one ADL compared to ex-occasional or never smokers.*

*Q. What role do the binary self-rated health variable (created in practical A3) and wealth quintiles (****wealth5****) variables play in this association?*

. mhodds ADL smok\_bin, by(newsrh)

Maximum likelihood estimate of the odds ratio

Comparing smok\_bin==1 vs. smok\_bin==0

by newsrh

--------------------------------------------------------------------------

newsrh | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------

VG or go | 1.346168 3.85 0.0497 0.99935 1.81334

fair or |1.083832 0.38 0.5391 0.83821 1.40142

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for newsrh

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.189616 3.07 0.0795 0.979530 1.444760

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1) = 1.17

Pr>chi2 = 0.2800

*The results of the test for homogeneity suggests that the stratum specific estimates are similar and therefore we look at the Mantel-Haenszel pooled estimate in the middle of the output. After accounting for self-rated health, the association between smoking and ADLs has reduced to 1.19 (95% CI: 0.98, 1.44). We can therefore conclude that self-rated health is a confounder of this association. We now repeat this analysis taking into account wealth.*

. mhodds ADL smok\_bin, by(wealth5)

Maximum likelihood estimate of the odds ratio

Comparing smok\_bin==1 vs. smok\_bin==0

by wealth5

--------------------------------------------------------------------------

wealth5 | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------

1 | 0.902878 0.27 0.6044 0.61334 1.32911

2 | 0.999040 0.00 0.9963 0.66240 1.50676

3 | 1.956201 8.63 0.0033 1.23977 3.08663

4 | 1.092608 0.18 0.6745 0.72255 1.65218

5 | 1.291787 1.24 0.2659 0.82185 2.03044

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for wealth5

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.173880 2.83 0.0923 0.973805 1.415062

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(4) = 7.58

Pr>chi2 = 0.1081

*The results of the test for homogeneity suggest that the stratum specific estimates are similar and therefore we again look at the Mantel-Haenszel pooled estimate. After controlling for wealth quintiles, the association between smoking and ADLs has again reduced substantially from the crude estimate (OR=1.17, 95% CI: 0.97, 1.42). We again conclude that wealth is a confounder of the association between smoking and ADLs.*

*In a next step it makes sense to adjust or control for both confounders simultaneously.*

. mhodds ADL smok\_bin newsrh wealth5

Mantel-Haenszel estimate of the odds ratio

Comparing smok\_bin==1 vs. smok\_bin==0, controlling for newsrh and wealth5

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.089907 0.73 0.3934 0.894318 1.328272

----------------------------------------------------------------

*Here we can see that after controlling for both self-rated health and wealth there is no longer an association between smoking and ADLs (OR=1.09, 95% CI: 0.89, 1.33). Therefore we conclude that this association was confounded by self-rated health and wealth.*

***C2 Considering confounders and effect modifiers using mhodds***

In the previous section we considered two confounders simultaneously. In this section we will combine confounders and effect modifiers in our analysis.

*Q. Test the crude association between social class and CRP using the mhodds command. Use the binary social class variable* ***manual.*** *Create a binary CRP variable from the continuous* ***crp*** *with a cut-point of 4mg/l.*

*A.*

. codebook crp

--------------------------------------------------------------------------

crp Blood CRP level (mg/l)

--------------------------------------------------------------------------

type: numeric (double)

label: hscrp, but 211 nonmissing values are not labeled

range: [.2,96.4] units: .1

unique values: 211 missing .: 958/3,129

examples: 1

2.5

6.7

.

. recode crp min/3.99=0 4/max=1, gen(crpbin)

. label define bicrp 0 "<4mg/l" 1 ">=4mg/l"

. label values crpbin bicr

. tab crpbin

RECODE of |

crp (Blood |

CRP level |

(mg/l)) | Freq. Percent Cum.

------------+-----------------------------------

<4mg/l | 1,593 73.38 73.38

>=4mg/l | 578 26.62 100.00

------------+-----------------------------------

Total | 2,171 100.00

. tab manual

Manual vs |

non-manual |

sclass | Freq. Percent Cum.

------------+-----------------------------------

non man | 1,844 60.26 60.26

manual | 1,216 39.74 100.00

------------+-----------------------------------

Total | 3,060 100.00

*26.6% of our sample have a CRP level above 4mg/l and 1216 (39.7%) were from a manual social class. Next we test the association between social class and CRP:*

. mhodds crpbin manual

Maximum likelihood estimate of the odds ratio

Comparing manual==1 vs. manual==0

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.685381 27.69 0.0000 1.384511 2.051635

----------------------------------------------------------------

*People in a manual social class have 69% higher odds of having a high CRP level compared to those in non-manual social classes (OR=1.69, 95% CI: 1.38, 2.05).*

*Q. Assess the role of ADLs and smoking status (binary variables of each used in section C1* ***ADL smok\_bin****) in the association between binary social class and binary CRP. Are they confounders or effect modifiers? Build a ‘final model’ and interpret your findings.*

*A. We will use mhodds to assess whether ADL and smoking might be confounders or effect modifiers.*

. mhodds crpbin manual, by(ADL)

Maximum likelihood estimate of the odds ratio

Comparing manual==1 vs. manual==0

by ADL

-------------------------------------------------------------------------- ADL | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+--------------------------------------------------------------- No ADLs | 1.671404 19.84 0.0000 1.32996 2.10051

1+ ADLs | 1.180065 0.62 0.4310 0.78118 1.78262

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for ADL

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.537552 18.17 0.0000 1.259653 1.876761

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1) = 2.10

Pr>chi2 = 0.1470

*Looking at the test of homogeneity suggests that ADL is not an effect modifier and we can also see this looking at the stratum specific estimates at the top of the output – even though there only appears to be an association for those with no ADLs, the overlap in the 95% CIs is substantial. We therefore cannot be confident there is a statistical difference in the strata. We therefore view the Mantel-Haenszel pooled estimate which shows that after accounting for ADLs, those from manual social class have 54% higher odds of having a high CRP value compared to those from non-manual social classes (OR=1.54, 1.26, 1.88). This estimate is somewhat lower than our crude estimate and therefore we consider it as a confounder.*

. mhodds crpbin manual, by(smok\_bin)

Maximum likelihood estimate of the odds ratio

Comparing manual==1 vs. manual==0

by smok\_bin

--------------------------------------------------------------------------

smok\_bin | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------never/ex | 1.189656 1.05 0.3051 0.85331 1.65858

ex/curre | 1.892052 25.13 0.0000 1.46826 2.43816

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for smok\_bin

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.600049 21.53 0.0000 1.309528 1.955024

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1) = 4.76

Pr>chi2 = 0.0291

*The test of homogeneity of ORs at the bottom of the output suggests that there is a statistical difference in the strata of smoking status. If we look at the stratum specific estimates we see that for those who were never or ex-occasional smokers that there is no association between social class and CRP. However, for those who were ex- or current regular smokers, being from a manual social class was associated with a 90% increased odds in having a high CRP value (OR=1.90, 95% CI: 1.48, 2.45). We therefore consider smoking status to be an effect modifier.*

*Next we build our final model with both smoking as an effect modifier and ADLs as a confounder.*

. mhodds crpbin manual ADL, by(smok\_bin)

Mantel-Haenszel estimate of the odds ratio

Comparing manual==1 vs. manual==0, controlling for ADL

by smok\_bin

--------------------------------------------------------------------------

smok\_bin | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------never/ex | 1.047141 0.07 0.7916 0.74400 1.47380

ex/curre | 1.767139 19.54 0.0000 1.36815 2.28249

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for ADL and smok\_bin

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

1.466124 13.78 0.0002 1.196462 1.796563

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1) = 5.81

Pr>chi2 = 0.0159

*For those who were never or ex occasional smokers, social class was not associated with CRP (OR=1.05, 95% CI: 0.74, 1.47). For those who were ex- or current regular smokers, being from a manual social class was associated with higher CRP (OR=1.77, 95% CI: 1.37, 2.28). This is after controlling for ADLs.*

***Optional extra practice***

In practical sections A3 and B1 we saw that self-rated health partially confounded the association between physical activity and heart attack, and that sex was an effect modifier. The role of self-rated health and sex were tested in separate analyses. For extra practice consider both of these variables in one single analysis. What do you conclude?

. mhodds heart\_attack physact newsrh, by(sex)

Score test for trend of odds with physact

controlling for newsrh

by sex

(The Odds Ratio estimate is an approximation to the odds ratio

for a one unit increase in physact)

--------------------------------------------------------------------------

sex | Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------+---------------------------------------------------------------

Men | 0.858507 1.97 0.1606 0.69372 1.06243

Women | 0.613588 8.47 0.0036 0.44160 0.85256

--------------------------------------------------------------------------

Mantel-Haenszel estimate controlling for newsrh and sex

----------------------------------------------------------------

Odds Ratio chi2(1) P>chi2 [95% Conf. Interval]

----------------------------------------------------------------

0.777343 7.62 0.0058 0.650033 0.929587

----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1) = 2.82

Pr>chi2 = 0.0930

*Once we have adjusted for self-rated health, sex no longer appears to be an effect modifier of the association. If we look at the pooled estimate which shows the association between physical activity and heart attack adjusted for self-rated health and sex, the estimate is further away from OR=1 (no effect) and closer to the crude estimate than when adjusting for self-rated health alone. In this case we would probably discard sex from our analyses and adjusted for self-rated health as a confounder.*