## Stata tips #4

***Epidemiologic Methods***

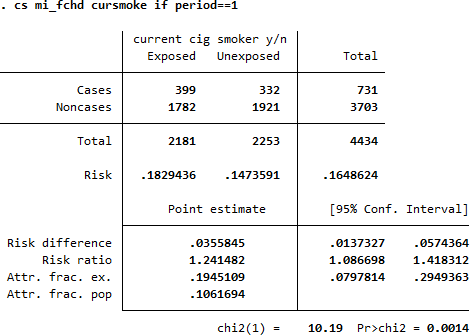
## Adapted from notes of previous TAs

This week:

1. Using cs, command (REVIEW)
2. Steps to identify confounding and effect modification using CS command in STATA

*Note: For this Stata session, we will be using the Framingham dataset again.*

# Cohort study command (cs)

The *cs* command can be run with two binary variables and this command is really just an extension of the tab command. We’ll analyze a couple of variables that are already in binary form by comparing *mi\_fchd1* and *cursmoke1.*

# Confounding (some basic discussion):

It is important to have a conceptual model/ framework before you throw everything in a model and control for confounders. It is important to note that if a variable is in the causal pathway of your conceptual model, then it is not a confounder. For example, if you were interested in studying Lipitor (a lipid lowering medication) and its effect on cardiovascular disease in a placebo controlled trial you should be careful regarding whether you control for “change in LDL.” In this scenario, you would only control for change in LDL if you were interested in the effects of Lipitor independent of the drop in LDL level.

In general, **confounding is identified by comparing the crude and adjusted odds ratios, relative risks**, etc. and determining if they are substantively different. There are multiple ways to do this. Some recommend using a **10% or greater difference between the crude and adjusted values** to call confounding, but it really is a judgment call based on subject knowledge regarding what constitutes a significant difference. It is important to note that

statistical significance does not play a role here: it is possible for a small study to show large differences between crude and adjusted estimates that are not statistically significant. This is still confounding. The opposite can also occur: a huge study where the effects are very similar but the 95% CIs don’t overlap at all. Here, this wouldn’t be confounding if the differences are substantively small. A corollary to this for the future is

that if, for example, subject matter knowledge strongly suggests age is a known confounder in the relationship of interest, don’t drop it from your statistical model if it’s not statistically significant!

# Effect modification (some basic discussion):

Effect modification occurs when the relationship between two variables of interest depends on the level of a **third variable**. For example, sex is a known effect modifier of the relationship between aspirin use and cardiovascular events: it is protective for men and not protective for women. What you will see in your data is that the relative risk or odds ratio of a relationship of interest will be different at different levels of the third variable, and the test of homogeneity will be statistically significant.

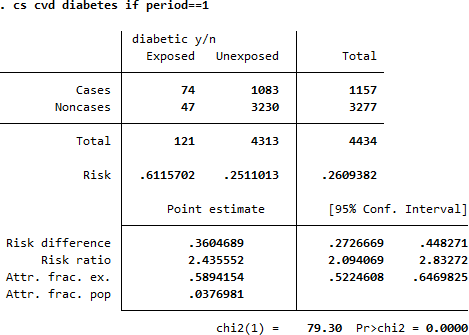
There are multiple ways to **test for confounding/effect modification.**

# Stratification

1. **Regression**

Today we will go over the steps for **identifying confounders using stratification**, which is an intuitive way to see how the numbers work out. In later classes we will learn about **using regression** to perform similar tasks.

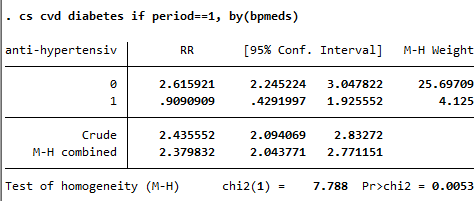
Let’s say we are interested in the relationship between diabetes and cardiovascular disease (cvd1). In this dataset, we can use the cs command as above, as this is a cohort study and we are interested in relative risks:



How would you **interpret the risk ratio here?**

**This unadjusted (crude) analysis shows that there is a statistically significant 2.4 times the risk of cvd in those with diabetes. (p < 0.001).**

Now let’s say we wanted to control for whether the subject was on BP meds.



## Is there confounding here?

These results are consistent with a quite similar increase in the risk of cvd with diabetes when we then take blood pressure medication use into account (MH RR estimate controlling for bpmeds1 = 2.38, 95% CI 2.04, 2.77) compared to the crude RR estimate from the analysis above of 2.44 (95% CI 2.09, 2.83). Given the similarity between the crude and adjusted estimates, we would conclude that use of blood pressure medications **didn’t confound** the association between diabetes and cvd.

***What about effect modification?*** i.e. is the association between diabetes and cvd the same for those on blood pressure medication versus those who aren’t? That is, is there any reason to suspect that blood pressure medications **modify** the association between diabetes and cvd? What do the stratified estimates tell us? It looks as if diabetes is positively associated with cvd among those who don’t use blood pressure medications (OR = 2.62, 95% CI 2.25, 3.05), but that there is no association between diabetes and cvd among individuals who do use blood pressure medication (OR = 0.91, 95% CI 0.43, 1.93).

The confidence interval on that latter estimate is quite wide, consistent both with a protective and a risk enhancing effect of diabetes on cvd. What does the test of homogeneity tell us in this case? The chi2 value is 7.788 and its corresponding p-value is 0.0053. This is strong evidence that we should reject the null hypothesis that the effect of diabetes on cvd is the same across the levels of the bpmeds1 variable. That is, there is statistical evidence for **effect modification**.

What if were interested in smoking as a potential confounder?

# . cs cvd1 diabetes1, by (cursmoke1)

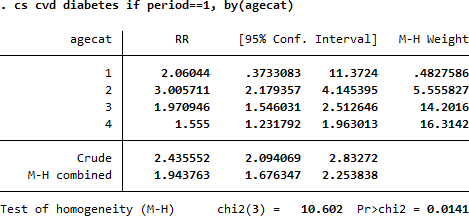
Is there confounding here? Effect modification? What about age?

How agecat was generated:

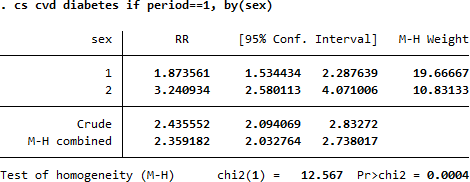
gen agecat=1

replace agecat=2 if age>40 replace agecat=3 if age>50 replace agecat=4 if age>60

Interpretation?

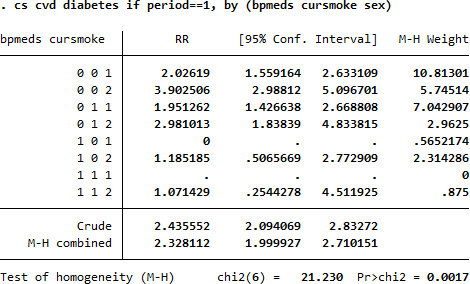


What about sex as a confounder/effect modifier?



Interpretation?

You can control for multiple variables if you want:



**Be careful** though, the more variables you add the more likely you will find a combination where no events occur.

HW #4:

| **Weight Category** | **Number of subjects**  **(Total)** | **Number of deaths** | ***Cumulative Incidence (Risk)*** | ***Relative Risk*** | **95% Confidence Interval of RR** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| **Underweight** | 167 | 68 | 0.407 | 1.51 | (1.25, 1.82) | p<0.001 |
| **Normal (ref.)** | 5032 | 1358 | 0.270 (ref.) | 1.00 (ref.) | (0.94, 1.07) | p<0.001 |
| **Overweight** | 4816 | 1511 | 0.314 | 1.16 | (1.09, 1.24) | p<0.001 |
| **Obese** | 1560 | 558 | 0.358 | 1.33 | (1.22, 1.44) | p<0.001 |

| **Weight Category** | **Number of subjects**  **(Total)** | **Number of deaths** | ***Incidence Density (Rate)*** | ***Incidence Rate Ratios*** | **95% Confidence Interval of RR** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| **Underweight** | 167 | 68 | 0.000054 | 1.59 | (1.23, 2.03) | p<0.001 |
| **Normal (ref.)** | 5032 | 1358 | 0.000034 | 1.00 (ref.) | (0.93, 1.08) | p<0.001 |
| **Overweight** | 4816 | 1511 | 0.00004 | 1.18 | (1.09, 1.27) | p<0.001 |
| **Obese** | 1560 | 558 | 0.0000465 | 1.37 | (1.24, 1.51) | p<0.001 |

| **Evaluting Obesity on Mortality:**  **Stratified Analysis Summary for Age Category, Sex and Current Smoking Status** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Risk Ratios** | | |  |  |  |  |  |
| **Crude RR / M-H RR for Obesity and Death** | **Age Category (30s) Stratum 1** | **Age Category (40s) Stratum 2** | **Age Category (50s) Stratum 3** | **Age Category (60+) Stratum 4** | **Counfounding?** | **Effect Modification?** | **p Homoge-neity** |
| 1.33 / **1.24** | 1.34 | 1.25 | 1.28 | 1.21 | No | No | p=0.907 |
|  | **Sex**  **(Male) Stratum 1** | **Sex (Female) Stratum 2** |  |  |  |  |  |
| 1.33 / **1.30** | 0.99 | 1.66 |  |  | No | **Yes** | **p<0.001** |
|  | **Current Smoking (No) Stratum 1** | **Current Smoking (Yes) Stratum 2** |  |  |  |  |  |
| 1.33 / **1.35** | 1.30 | 1.43 |  |  | No | No | p=0.251 |