CONFOUNDING

SLIDES PREPARED BY JENNIFER D. PECK, PHD
AND OTHER BSE FACULTY

Welcome to this presentation focused on confounding.

Confounding

Learning Objectives:

- Identify and describe methods available to control for confounding
- Use stratification to evaluate potential confounders

2

Our learning objectives for this session are to identify and describe methods available to control for confounding. In particular, we will focus on the use of stratification to evaluate potential confounders.

Interpretation of Study Results

Results may be due to:
- Causal association
- Chance (random error)
- Bias (systematic error)
- Confounding

Let's begin by discussing considerations for the interpretation of study results. So when we are assessing the association between an exposure and a disease when we observe an association or even a lack of association we have to consider possible explanations for the observed results. And so there are always four things that you should consider. The first would be that the exposure is causally related to the outcome. This is typically our motivation for conducting the study. We want to identify causal factors for disease So one possibility for the association assuming we have observed an association one possibility is that that association is causal. But we can't jump to that conclusion without first ruling out other possible explanations. One possible explanation that we should always consider is the role of chance the results we have observed could be due to random error. Alternatively if there is a systematic error in the way that we have conducted our study our results could be due to bias and additionally the results could be due to confounding. So these are the four possible explanations that you should consider causality, chance, bias and confounding. And now we want to move on to talk a little more about what we mean specifically by bias and confounding.

Confounding

- A distortion of the exposure-disease association due to the influence of a third factor
- A confounder may fully or partially account for the observed effect of the study exposure or mask or hide an underlying true association
- Results when another factor is unevenly distributed between comparison groups and may account for the observed association

4

Confounding is defined as a distortion of the exposure-disease association due to the influence of a third factor.

A confounder may fully or partially account for the observed effect of the study exposure or mask or hide an underlying true association

Confounding results when another factor is unevenly distributed between comparison groups and may account for the observed association

Criteria for a Confounding Variable

1. It must be a risk factor for the disease, or associated with the disease, but not necessarily causal.

2. It must be associated with the exposure under study, but is not a result of it.

3. It must not be an intermediate variable on the causal pathway

5

The conventional approach to assessing confounding indicates that a confounder should meet three criteria. First it should be associated with the disease independent of exposure. So this means there should be a known association between the disease and the potential confounders of interest and this relationship should exist among the non-exposed. Second the factor should be associated with exposure in the source population and third the factor must not be an intermediate variable on the causal pathway which relates back to the phrase in parentheses on item two and that in relation to the exposure we don't want our potential confounders to be a result of exposure and therefore that exposure is only influencing the outcome through this intermediate or potential confounding factor. So if a factor lies on the causal pathway it does not meet the criteria for confounding. We will walk through a few examples to demonstrate this on the next few slides.

Interrelationship Between Confounder, Exposure and Disease

Exposure → Disease
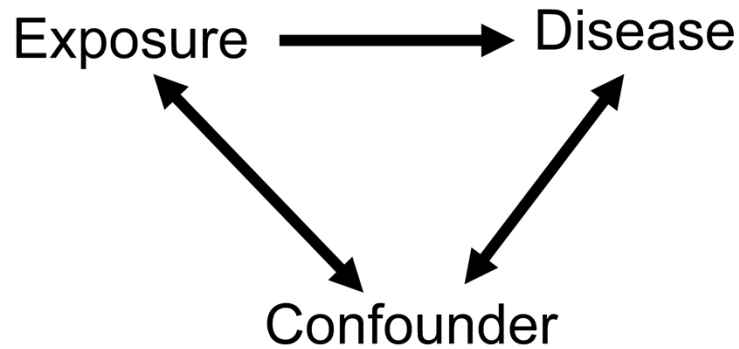
Confounder

6

This slide depicts the classic confounding triangle that demonstrates the central relationship between the confounder and the exposure and the disease. So in general terms this represents the relationships that we discussed in the previous slide where the confounder must be associated with both the exposure and the disease and of course in addition to that not be an intermediate factor on the causal pathway.

So if you consider this diagram then if there is no association between the exposure and the potential confounder, in other words if you were to block this path, then there can be no confounding by that factor.

Similarly, if the confounding factor has no relationship with disease, so you have blocked the path between the confounder and the disease or the outcome. Then there can be no confounding by that factor.

In the next slide we want to walk through a specific example using the relationship between alcohol consumption and myocardial infarction.

A Confounder Cannot Be An Intermediate Link on Causal Pathway

Exposure → Confounder → Disease

Alcohol → ↑ HDL Levels → ↓ MI
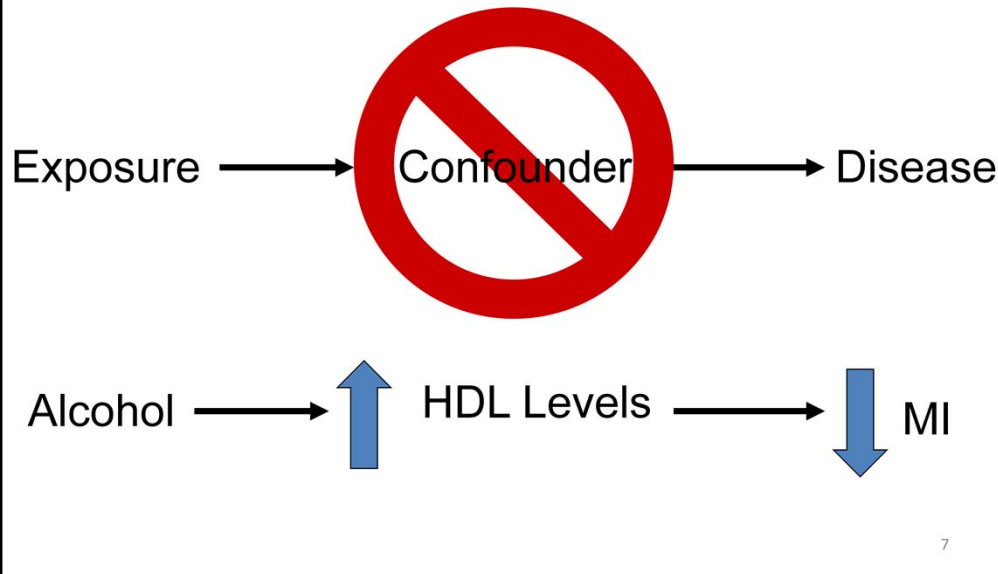
In this case, HDL is associated with alcohol consumption and is associated with MI; however, HDL is increased as a result of alcohol consumption, so, HDL meets criteria 1 and 2, but not 3. Meaning, HDL is associated with disease, associated with exposure, but is in the causal pathway. Therefore, HDL is not considered a confounding factor but instead may be a mediating factor.

In order to determine whether a factor lies on the causal pathway we have to rely on existing knowledge including knowledge of the biological mechanism that is underlying this relationship between exposure and disease.

If the mechanism of action of the exposure is to alter levels of the potential confounder which would then in turn influence disease, then that factor would not meet our criteria for confounding but instead would be considered an intermediate step on that causal pathway, that causal chain of events, that are occurring between exposure and disease.

So here in the example of a study of the effect of moderate alcohol consumption on decreasing risk of myocardial infarction. At first glance the levels of high density lipid protein cholesterol might it might appear to meet the criteria for confounding because it appears that HDL levels are associated with both alcohol use and the risk of myocardial infarction because we know that increased alcohol use raises HDL levels. The effect of alcohol on the risk of myocardial infarction is acting through HDL levels. So if this is the mechanism of action, then HDL would not be considered to meet our criteria for confounding and therefore we would not want to attempt to control for this factor in our

analyses.

Example of Age as a Confounder

Moderate Alcohol	MI Status	
	Cases	Controls
Exposed	30	18
Not Exposed	70	82
	100	100

Odds Ratio = (30 x 82) / (18 x 70) = 1.95

Is the association observed a causal one, or could it have resulted from differences in the age distribution of cases and controls? In other words, is the association confounded by age?

Now, let's continue with our example investigating the association between moderate alcohol consumption and MI.

In this example, the odds of moderate alcohol consumption for those with an MI are 95% higher than the odds of moderate alcohol consumption for those without a history of MI.

Before drawing inference from these data, we would want to consider other possible factors that may be confounding this association, such as age.

Criterion 1:
Is age related to the disease?

	Cases	Controls
< 40 years	50	80
> 40 years	50	20
	100	100

Older age (> 40 years) is associated with being a case, and younger age with being a control.

As we begin assessing whether age may be a confounder of this relationship, we first ask whether or not age is related to case status.

As shown in the previous slide we have one hundred cases and one hundred controls so the counts that are in these cells also represent percentages. We see that 80% of the controls are in the younger age category as compared only 50% of cases and so we can conclude from this that we do see a difference in the age distribution between cases and controls and we would conclude that age is associated with disease.

Criterion 2:
Is age related to exposure?

Age	Total	Exposed	Not exposed	% Exposed
< 40 yrs	130	13	117	10
≥ 40 yrs	70	35	35	50

Older age is associated with a higher prevalence of exposure.

10

Next we would evaluate whether age is related to exposure.

We see that 50% of those who are in the older age category are exposed compared to only 10% of those in the younger age category. From these data we would conclude that older age is associated with a higher prevalence of exposure.

Our answer to this question would be yes, age does appear to be related to exposure as well which would make the second criterion for confounding.

So given these data, age meets the criteria for confounding. We have established that age was associated with the disease; we saw a higher distribution of controls who were in the younger age category. Second, age is associated with the exposure, but as far as we know it is not a result of it and so we are also assuming that age is not an intermediate on the causal pathway. So we would conclude that in this example age would meet the criteria for confounding.

Is association between exposure and disease causal, or is it confounded by age?

		Cases	Controls	OR
Age < 40	Exposed	5	8	
	Not Exposed	45	72	1.0
Age $\geq$ 40	Exposed	25	10	
	Not Exposed	25	10	1.0

So the steps we have taken so far have only assess whether or not age meets the criteria for confounding. But what we have not done is to actually assess whether there is evidence for confounding in this relationship.

The next step that we want to take in order to investigate age as a confounder is to conduct a stratified analysis. This means that we're breaking up our data into groupings or strata that are defined according to the levels of this potential confounding factor.

In this case, age is our potential confounding factor and so we have separated the data into two strata, those who are less than age forty and those who are greater than or equal to age forty.

We are setting up separate two by two tables and calculating separate odds ratios for each of the age subgroups.

Among those who are less than 40 we would calculate an odds ratio (comparing the odds of moderate alcohol consumption between MI cases and controls) of 1.0. For those who are greater than or equal to age 40, we would calculated odds ratio also 1.0.

After stratifying by age, we see a different association between alcohol and MI. In fact no association is evident with an odds ratio of 1.0, which indicates that there is no association between exposure and disease.

Conclusion

- **Stratified analysis reveals that within each age group there is no association between exposure and disease (OR = 1).**

- **The original OR of 1.95 was due to the difference in the age distribution between cases and controls.**

- **Thus, the association was confounded by age.**

12

The stratified odds ratio values of 1 are different from the crude estimate that we calculated a few slides back which was a ratio of 1.95.

By controlling for confounding, what would have originally appeared to be an association between the exposure and disease is now removed and that we have no evidence of association when we control for the effect of age on that exposure/disease association.

Steps for Evaluating Confounding

1) Calculate crude overall estimation of the exposure-disease association

2) Stratify the data by levels of the suspected confounder

3) Calculate the stratum-specific estimates

4) Compare the stratum-specific estimates to one another and to the crude measure of association

13

In order to assess confounding we need to follow a few steps and this slide just provides a reference that basically outlines the steps we just followed.

First, when evaluating confounding we calculate the crude overall estimation of the association between exposure and disease.

Next, we stratify the data by meaningful levels of the suspected confounder. In our example, the potential confounding factor was age.

Then you would calculate the stratum-specific estimates, so you're calculating the association between exposure and disease within subgroups of participants. In our example, the subgroups would be younger and older age groups.

Finally, we would take the results from the stratified analyses and compare them to the crude measure of association to determine if they are similar or if they are different.

If they are the same, we would conclude there is no evidence for confounding and if they differ substantially, we would conclude that

confounding is evident. If there is evidence of confounding, we would need to control for that factor in order to have an unbiased estimate.

While there is no black and white criteria to use when comparing the crude to the stratified estimates when detecting confounding, there are general guidelines of ten or twenty percent differences that are used as evidence of confounding.

Example of Confounding

	Males		Females		Total	
	MI	No MI	MI	No MI	MI	No MI
Smokers	21	17	18	7	39	24
Non-Smokers	26	59	88	95	114	154
	OR = 2.8		OR = 2.8		OR = 2.2	

Crude OR = (39) 154 / 24 (114) = 2.2
OR for Males = (21) 59 / 17 (26) = 2.8
OR for Females = (18) 95/ 7 (88) = 2.8

14

Let's consider another numeric example assessing the association between smoking status and myocardial infarction. We are concerned that gender may be acting as a confounding factor.

So to follow the steps that we outlined in a previous slide we would begin by estimating the crude association between smoking and myocardial infarction.

So in this case, the study is a case -control study and so we're calculating an odds ratio to quantify the association between exposure and outcome.

The overall crude odds ratio is 2.2. After stratifying by gender, we see that the odds ratio for males is 2.8 and the odds ratio for females is 2.8. This is a modest increase but potentially meaningfully different from the crude association.

Based on these data we would conclude that the association between smoking and myocardial infarction is confounded by gender.

Direction of Confounding

- Positive Confounding (Away from the Null)
  - The confounding factor produces an observed estimate of the association between exposure and disease that is an <u>overestimate</u> – either more positive or more negative – than the true association.

- Negative Confounding (Toward the Null)
  - The confounding factor produces an observed estimate of the association between exposure and disease that is an <u>underestimate</u> of the true association.

15

We can describe the influence of confounding as being either in the positive or negative direction.

Positive confounding would cause the exposure/disease association to be an overestimate of the true association whereas a negative confounding would cause that observed association to be an underestimate of the true association.

So when we're speaking of underestimates, the observed effect or the crude odds ratio would be closer to 1.0 (either above or below 1). We could call that leaning toward the null, whereas a positive confounder would be described as producing an effect that is away from the null.

Direction of Confounding

Example	Crude OR	Stratum 1 OR	Stratum 2 OR	Direction?
1	2.0	1.2	1.2	Positive
2	0.5	0.85	0.85	Positive
3	2.0	3.2	3.2	Negative
4	0.5	0.35	0.35	Negative

16

So here's an opportunity to practice your skills trying to interpret the direction of confounding. So here are four examples. Hypothetically we're using four different data sets here and you are given the crude association in the first column and the stratum-specific estimates in the following two columns.

The first example is an indication of positive confounding because the crude OR is 2 while the stratum-specific estimates are 1.2 and are closer to the null value of 1.

The second example is also an indication of positive confounding because the crude OR is 0.5 while the stratum-specific estimates are 0.85 and are closer to the null value of 1.

The third example is an indication of negative confounding because the crude OR is 2.0 while the stratum-specific estimates are 3.2 and are farther from the null value of 1.

The fourth example is an indication of negative confounding because the crude OR is 0.5 while the stratum-specific estimates are 0.35 and are farther from the null value of 1.

Avoiding and Controlling for Confounding

- The aim is to control confounding and <u>eliminate its effects</u>
- Methods for controlling confounding
  - Study Design
  - Data Analysis

17

We have discussed the fact that confounding is a distortion of the true association and so our goal then is to control for it.

We do not want confounded estimates, so we want to eliminate the effects of confounders and produce estimates that are not biased by the impact of the confounding factor.

There are several methods that we can use to control for the confounding factor and those methods could be utilized within the study design or at the stage of the data analysis. We will now discuss several methods that can be used to control for confounding.

Avoiding and Controlling for Confounding

- In designing and conducting study:
    1. Randomization
    2. Restriction
    3. Matching
- In analyzing the data:
    1. Stratification
    2. Multivariate analyses

18

Within the design of the study, there are three methods that you could use including, randomization, restriction, and matching. And then when analyzing the data, you can utilize either stratification, which you've already been introduced to, or multivariate analysis.

We will now discuss each of these approaches.

Randomization

- Randomly assigning exposure to ensure equal distribution of confounders in each exposure category
- Advantages
  - Provides control of known and unknown factors
  - Convenient, inexpensive, straightforward data analysis
- Disadvantages
  - Can be applied only to intervention studies
  - Works well only for large sample sizes

19

We have discussed randomization in the context of clinical trials or experimental designs.

With randomization, you are randomly assigning exposures in order to ensure that you have the equal distribution of confounders in each exposure category.

So by assigning a treatment to participants, you are controlling who is getting the treatment assignment and who is not and effectively controlling for confounding.

An advantage is that this provides control for both known and unknown confounders.

In the context of an experimental design it is convenient, straightforward, and inexpensive to use randomization.

However the major disadvantage is that this method simply is not applicable to observational studies and we know that many questions that are relevant to public health cannot be addressed in an experimental design for ethical

purposes, particularly when we are interested in risk factors for disease.

So if it's a question that's beyond prevention or intervention, then we need to rely on methods that can be utilized in an observational design.

Another point to remember is that randomization will work best for a large sample sizes. So if the number of subjects in your study are small, there's still a chance that remains that the distribution of these confounding factors will still be different across the study group.

Restriction

- Limiting inclusion criteria to prohibit variation
  - Example: Restricting subjects to those within a narrow age range
- Advantages
  - Effective in providing complete control of known confounders
  - Convenient, inexpensive, straightforward data analysis
- Disadvantages
  - Shrinks pool of available study subjects
  - Residual confounding if restriction not sufficiently narrow
  - Does not permit evaluation of E→D across levels of the restricted factor
  - May limit generalizability

20

The method of restriction is used when we limit the study to a particular subgroup.

So we've limited our inclusion criteria in such a way that we have removed any source of variation across that variable.

For example, we could restrict our studies simply to females and then there's no variation by sex and hence, sex cannot act as a confounder because exposure groups are balanced by sex.

As another example, we can restrict our study to those within a narrow age range and so there's less variation by age and age will then not function as a as a confounder.

Remember that confounding can only occur if there's an unequal distribution of that factor across your comparison groups and so if you hold that constant across everyone, there's no variation, there's no difference in the distribution, and that factor cannot function as a confounder.

Restriction is a relatively straightforward method to utilize and it's very effective in providing complete control for known confounding factors that you

can identify and then utilize in your study design to restrict your study population.

There is no added expense and restriction is straight forward.

However there are a number of disadvantages to using restriction. For one, it would shrink the pool of available study subjects. For example, by restricting your study to females, you've eliminated half of the potentially eligible population.

You then have fewer subjects available then to participate in the study. Also you can still have what we refer to as residual confounding. If that restriction is not sufficiently narrow, meaning for instance if we were to restrict by age but we still had a relatively wide age range that was eligible to participate in the study, you could still have confounding left over or age could still function as a confounder. For example, if you allowed anyone under the age of 40 to participate in your study then if that you know age met the criteria for confounding age could still function as a confounder because there's a wide age range there even if you limited your study to adults.

If you limited your study perhaps to a 5-year age range or a 2-year age range, age would have less influence as a confounder.

The other disadvantage of restriction is that when you utilize restriction as a method to control for confounding, you cannot evaluate the exposure/disease association across all levels of that restricted factor because you don't have those other levels available to you. As an example, you couldn't assess the association between smoking and myocardial infarction to determine how that may differ in males and females if you've already restricted your study to females alone.

If the impact of gender on the association was a question of interest, you would not want to restrict your sample by gender. Also note that restriction may limit the generalizability of the study results. If we have a study that was restricted to white males, the results may or may not be generalizable to all human males but certainly would not be generalizable to females and so we have to keep this in mind when we apply restriction criteria to our studies to address confounding.

Matching

- Selecting subjects according to the value of suspected confounders to ensure equal distributions among study groups
- Advantages
  - Smaller sample size requirements for cohort studies
  - Useful when there would not be a sufficient number of subjects alike to control for these factors in the analysis
  - When case series is small, allows enough subjects in each strata to control for these factors in the analysis
  - When number of cases is small, R:1 matching can increase statistical power
- Disadvantages
  - Can be costly and time-consuming, requiring extensive searching and recordkeeping to find matches
  - May introduce confounding in case-control studies
  - Matching factor can no longer be evaluated as a risk factor

21

Matching is another approach to limit the impact of confounding factors. In this case, we select subjects according to the value of the suspected confounder to ensure equal distributions between study groups.

As an example, we may be concerned that gender will confound the association between cigarette smoking and MI. In a case-control study, we would match the MI cases to the MI controls using gender as the matching variable. A control, or multiple controls, would be selected for each case to have the same gender.

There are several advantages to matching. First, matching may result in smaller required sample sizes for cohort studies. Matching is useful when the confounding factor cannot be controlled for in the data analysis due to sample size issues. Also, multiple control participants can be matched to a single case to ensure that the confounding factor is controlled and would also result in increased statistical power.

Matching is not without disadvantages. Matching at the sampling stage can be costly and time-consuming and may require extensive searching and recordkeeping in order to find matches. Also, matching may introduce confounding in case-control studies depending on how the matching factors are defined. Finally, the matching factors, because they are balanced between diseased and non-diseased participants, can no longer be evaluated as a risk factor.

Stratification

- Evaluating the association between exposure and disease within homogeneous categories (strata) of the confounding variable
- Advantages
  - Allows for clear understanding of interrelationship among exposure, disease and confounding variables
  - Direct and logical strategy; computations easy to carry out
  - Minimal assumptions required for analyses to be appropriate
  - Permits the evaluation of effect modification
- Disadvantage
  - Inability to control simultaneously for several confounders
  - Continuous variables must be categorized for analysis

22

Another method that can be used to control for confounding is stratification. This occurs at the analysis step of the study.

To implement stratification, we evaluate the association between exposure and disease within homogeneous categories (strata) of the confounding variable. For example, if we are concerned that gender confounds the association between smoking and MI, we could compare the crude odds ratio value to the odds ratio estimated among males and the odds ratio estimated among females to determine if the crude and stratified values differ, in which case, we would conclude that gender acted as a confounder.

Advantages of stratification include the fact that stratification allows for a clear understanding of the interrelationship among exposure, disease and confounding variables. Stratification is a direct and logical strategy and computations are easy to carry out. Stratification involves minimal assumptions for the analyses to be appropriate. Stratification also permits the evaluation of the potential confounding factor in subsequent analyses.

Disadvantages include the inability to control simultaneously for several confounders and continuous variables must be categorized before stratified analyses can be implemented. This categorization may result in a loss of important information.

Example of Stratification

	Males		Females		Total	
	MI	No MI	MI	No MI	MI	No MI
Smokers	21	17	18	7	39	24
Non-Smokers	26	59	88	95	114	154
	OR = 2.8		OR = 2.8		OR = 2.2	

Crude OR = (39) 154 / 24 (114) = 2.2
OR for Males = (21) 59 / 17 (26) = 2.8
OR for Females = (18) 95/ 7 (88) = 2.8

23

We previously considered this data table as an example of a stratified analysis. In this case, we have stratified the data according to sex. We compare our stratum-specific estimates to the crude value to determine whether or not there is evidence for confounding. In this case, the OR for males and females differs from the crude OR and therefore, we conclude that gender acts as a confounding factor.

Multivariate Modeling

- Constructing mathematical models to describe the association between exposure and outcome variables
- Advantages
  - Allows for control of a number of confounding factors simultaneously
  - Useful in situations where stratification would fail because of insufficient numbers
- Disadvantages
  - The choice of the appropriate model is complex
  - Efficient mathematical modeling can often occur at the expense of a clear understanding of the data
  - Great potential for misuse

24

The remaining technique that we want to discuss at the data analysis stage is multivariate modeling.

This is the application of mathematical models to describe the association between the exposure and the outcome variable where we use mathematical modeling methods to adjust for multiple confounding factors.

There are several advantages of multivariate modeling. With modeling, we are able to control for more than one potential confounding factor simultaneously and so this is an attractive feature of the multivariate modeling.

Multivariate modeling is also useful in situations where stratified analyses would otherwise fail because you had insufficient numbers in certain cells or zeros in certain cells.

Disadvantages include the fact that choosing the appropriate model is quite complex. There can be complex relationships among exposures, disease outcomes and confounders and modeling decisions and interpretation of the results can be difficult. There is a potential for misuse of these modeling techniques when one is not familiar with whether or not modeling assumptions are being appropriately met.

Coffee, Caffeine, and Risk of Depression Among Women

Michel Lucas, PhD, RD; Fariba Mirzaei, MD, MPH, ScD;
An Pan, PhD; Olivia I. Okereke, MD, SM;
Walter C. Willett, MD, DrPH; E´ilis J. O'Reilly, ScD;
Karestan Koenen, PhD; Alberto Ascherio, MD, DrPH

25

Now, let's consider an example from a published cohort study investigating the association between coffee consumption and the incidence of depression.

Results: Table 2

Table 2. Relative Risks (95% CIs) of Clinical Depression According to Caffeinated Coffee Consumption[a]

			Caffeinated Coffee Consumption			
Variable	≤1 Cup/wk	2-6 Cups/wk	1 Cup/d	2-3 Cups/d	≥4 Cups/d	P for Trend
No. of cases	670	373	905	564	95	…
Person-years	113 305	63 322	163 256	105 619	17 960	…
Age-adjusted model[b]	1 [Reference]	1.02 (0.90-1.16)	0.96 (0.87-1.06)	0.90 (0.80-1.01)	0.88 (0.71-1.09)	.03
Multivariate model[c]	1 [Reference]	1.00 (0.88-1.14)	0.92 (0.83-1.02)	0.85 (0.75-0.95)	0.80 (0.64-0.99)	<.001
Sensitivity model[d]	1 [Reference]	0.90 (0.84-1.01)	0.96 (0.86-1.09)	0.88 (0.78-0.98)	0.82 (0.66-0.98)	.005

26

In this case, we see that incidence is calculated as the number of cases relative to the person-years at risk.

Regression models were used to adjust the association between coffee consumption and depression for confounding factors such as age, total energy intake; current menopausal hormones; smoking status; body mass index; physical activities; marital status; not involved in church, volunteer, or community group; retired; reported diagnosis of diabetes mellitus, cancer, high blood pressure, or myocardial infarction or angina; and Mental Health Index score.

We see that with adjustment, the relative risk estimates are shifted away from 1, hence, the protective effect of coffee consumption is greater in magnitude with adjustment.

Selection of Potential Confounders

- Selection is based on:
  - Existing knowledge
  - Evaluation of the data
  - Best judgment of the investigator

- Selection is **not** based on statistical significance!

27

We've discussed the definition of confounding and how to assess confounding, but how do you identify factors that might be potential confounders?

The selection of potential confounders is reliant on background knowledge of the disease and knowledge gained from existing literature where similar study questions have been evaluated. We can use our data to some extent to determine which measured variables may be acting as confounders. But, ultimately we have to know long before we collect the data which variables may act as confounders. It's important to give careful thought to the selection of potential confounders at the design stage of the study.

One point that we do want to emphasize is that we do not select confounders by assessing the statistical significance of the association between the confounding factor and the exposure or the confounding factor and the disease. There is no statistical test that will tell you whether or not something is confounded or whether or not it's functioning as a confounder.

For example, in a small study, a confounding factor could influence the magnitude of the association between exposure and disease even though it does not achieve statistical significance in its own relationship with either exposure or disease simply because there is not enough power to detect an

association that did exist.

Similarly, you can have a very large sample size where almost all associations are statistically significant even though the factor is not acting as a confounder.

Statistical significance is not a criterion on which to base a decision that a particular factor is or is not a potential confounder. We use our knowledge of the content area and existing evidence to determine whether or not a factor meets our criteria for confounding.

Test your knowledge of the key characteristics of a confounding variable…

Now, let's consider some review questions.

Is the variable in question a confounder?

- A study of the relationship between contact lens use and risk of eye ulcers.
- Crude RR = 3.0
- Age-Adjusted RR = 1.5

Is the variable in question a confounder?

A study was conducted to determine if contact lens use was a risk factor for eye ulcers.

The crude, unadjusted relative risk was 3.0 while the age-adjusted relative risk was 1.5.

The adjusted relative risk demonstrates a 50% change from the crude relative risk and therefore, age is acting as a confounder. In this case, confounding is in the positive direction because the crude estimate is farther from the null value of 1 than the adjusted relative risk.

Is the variable in question a
confounder in this study?

- A study of the relationship between exercise
and heart attacks conducted in men who do
not smoke. Is gender a confounder?

30

In a second example, a study of the relationship between exercise and heart attacks conducted in men who do not smoke.

Is gender a confounder?

No, gender cannot act as a confounder because gender is constant in the study. Through restriction, gender was limited to men only and therefore, gender is not associated with the exposure, exercise, or the outcome, heart attacks, and cannot act as a confounder.

Identify method for controlling for a confounder

- A study of exercise and myocardial infarction that includes men and women. Study determines relative risk separately for men and women and compares these to crude relative risk.

31

In this example, identify the method that was used to control for confounding.

A study of exercise and myocardial infarction that includes men and women was conducted. The investigators determine a relative risk separately for men and women and compare these to the crude relative risk.

Which method was used to account for confounding?

In this case, stratification was used because the stratified estimates are compared to the crude estimate.

Identify method for controlling for a confounder

- A case-control study of exercise and myocardial infarction that includes men and women. Controls are selected so that the proportions of male and female subjects groups are identical.

In another example, a case-control study of exercise and myocardial infarction was conducted that included men and women. Controls were selected so that the proportions of male and female subjects groups in the controls are identical to the proportion selected among the cases.

This is an example of frequency matching. The cases and controls are balanced by gender through matching.

Summary

- Identified methods available to control for confounding
  - Design
  - Analysis
- Use stratification to evaluate potential confounders

33

We discussed multiple methods to control for confounding at the design and data analysis stages. Methods at the design stage include restriction, matching, and randomization. At the analysis stage, we can use multivariate modeling or stratification to account for confounding. In this session, we demonstrated the use of stratification to evaluate the impact of confounding factors.

This concludes the presentation for confounding.