What is Confounding?

Suppose that a study finds that coffee drinking is associated with the risk of cancer of the pancreas.

Should we all stop drinking coffee? Maybe not. First we must make sure there is no obvious alternative explanation.

Perhaps coffee drinkers (many of whom smoke cigarettes) are at risk of cancer of the pancreas because they smoke, not because they drink coffee. If so, we could say that the apparent association of coffee drinking with cancer of the pancreas is due to **confounding** by cigarette smoking.

Confounding is the situation where an association between an exposure and an outcome is entirely or partially due to another exposure (called the confounder).

A variable will only confound an association if it satisfies three conditions:

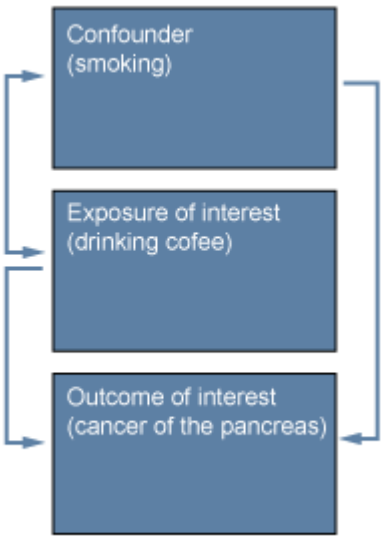
1. It must be associated with the exposure of interest ***among the source population*** (represented by the controls in a case-control study).

2. It must be a risk factor for the outcome of interest ***among the non-exposed***.

3. It must not be ***on the causal pathway*** between the exposure of interest and the outcome of interest.

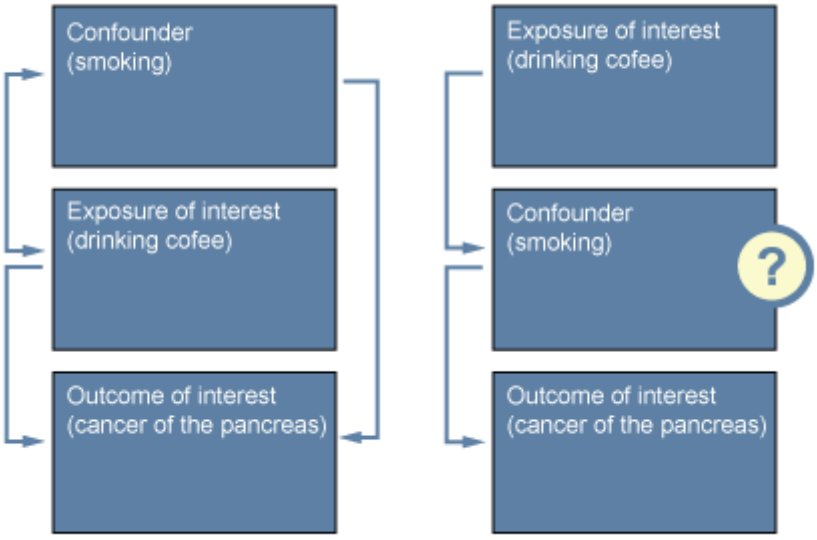
In practice when testing for confounding the following slightly looser definitions are used. A variable will only confound an association if it satisfies three conditions:

- - it must be associated with the exposure of interest
- - it must be a risk factor for the outcome of interest
- - it must not be on the causal pathway



To know whether the third condition is satisfied or not we have to consider whether coffee drinking ***causes*** a person to smoke more, which in turn ***causes*** pancreatic cancer. It is unlikely in this case that smoking is on the causal pathway because drinking coffee does not ***cause*** someone to smoke more. We will consider some more examples of this later in this session.

Smoking therefore fulfils all three conditions of a potential confounder for the association between drinking coffee and pancreatic cancer.



How does confounding occur?

Let's take our example of coffee drinking and cancer of the pancreas, and see how it works numerically.

Suppose that a case-control study finds that coffee consumption is associated with an increased risk of cancer of the pancreas; the basic data were as shown below.

What is the odds ratio for drinking coffee among cases (people with cancer of the pancreas) compared with controls?

	Cases (cancer of the pancreas)	**Controls**
Coffee	450	200
No Coffee	300	250

So the odds ratio is given by:

$$\frac{450 * 250}{300 * 200} = \mathbf{1.9}$$

If we now reanalyse the data, grouping the subjects according to smoking habit, we have the table below.

What are the odds ratios for drinking coffee among cases (people with cancer of the pancreas) compared with controls?

		Cases	Controls
Smokers	Coffee	400	100
	No Coffee	200	50
Non-smokers	Coffee	50	100
	No Coffee	100	200

Odds Ratio for drinking coffee among smokers =

$$OR = \frac{400*50}{100*200} = \mathbf{1}$$

Odds Ratio for drinking coffee among non-smokers =

$$OR = \frac{50*200}{100*100} = \mathbf{1}$$

In this reanalysis, when we look at smokers and non-smokers separately, there is no association between coffee drinking and cancer of the pancreas. So in this example smoking did indeed ***confound*** the association between coffee consumption and cancer of the pancreas.

A confounder must be associated with the outcome in its own right, not just because of its association with the exposure of interest.

The tables below show that smoking satisfies this condition for a confounder. It is associated with cancer of the pancreas (our outcome) even in people who do not drink coffee (the non-exposed).

	Cases	**Controls**
Coffee		
smokers	400	100
non-smokers	50	100
No Coffee		
smokers	200	50
non-smokers	100	200

Odds Ratio for smoking among coffee drinkers = (400*100)/(100*50)= 8

Odds Ratio for smoking among non coffee drinkers = (200*200)/(50*100) = 8

Still using the same example, let us see if smoking is associated with coffee drinking.

		Cases	**Controls**
Smokers	Coffee	400	100
	No Coffee	200	50
Non-smokers	Coffee	50	100
	No Coffee	100	200

Reorganizing the data:

	Coffee	**No Coffee**
Cases	400/450	200/300
% smokers	= 89%	= 67%
Controls	100/200	50/250
% smokers	= 50%	= 20%

So smoking is a potential confounder because it is associated with coffee drinking (the exposure of interest). This is true for both the cases and more importantly for the controls (which represent the source population).

Any specific risk factor can be an exposure of interest in one study, and a confounder in another - it all depends on what we are studying.

In the example, smoking was a confounder because the association that we were investigating was between coffee drinking and cancer of the pancreas.

If however we wanted to study the association between smoking and cancer of the pancreas, then smoking would be the exposure of interest, but we would also want to collect data on other exposures that might potentially confound this relationship, such as dietary fat intake, alcohol consumption, and perhaps also coffee consumption.

So a confounder is an exposure which explains (entirely or partly) an observed association between an exposure and an outcome, because it is associated with both the exposure of interest and the outcome.

Positive and Negative Confounding

In the first example we used, the effect of the confounding variable (cigarette smoking) was to cause an apparent association to be observed between an exposure (coffee) and an outcome (cancer of the pancreas), when in fact no association existed.

In this example, where the effect of a confounder is to make the observed association between exposure and outcome appear stronger (i.e. to increase the odds ratio) may be called **positive confounding**. Positive confounding results in a false positive association, or the exacerbation of a true association due to the distribution of a third variable.

This effect can also work in the opposite direction: confounding can also result in the association between an exposure and an outcome appearing to be weaker than it really is. This is called **negative confounding**. Confounding which masks a true association by making it appear smaller than it is in reality (false negative result).

This is illustrated by the following example:

In a study of the causes of diarrhoea in the UK, the data for the association between socioeconomic status (SES) and diarrhoea were as shown below.

	Diarrhoea	No Diarrhoea	Total
Low SES	159	341	500
High SES	92	408	500

The odds ratio for being in the lower socioeconomic group among people with diarrhoea was:

Odds Ratio = 2.1

This odds ratio of 2.1 suggests that there is an association between diarrhoea and social class. However, this estimate of the strength of the association may be confounded by other variables, and one such potential confounder is travel to a tropical country. Here are the same data, analysed separately for those who had travelled to a tropical country or not.

What is the odds ratio for being in the lower socioeconomic group among people with diarrhoea compared with people without diarrhoea in each travel category?

		Diarrhoea	No Diarrhoea	Total
Travel to Tropics	**Low SES**	15	5	20
	High SES	72	108	180
No Travel to Tropics	**Low SES**	144	336	480
	High SES	20	300	320

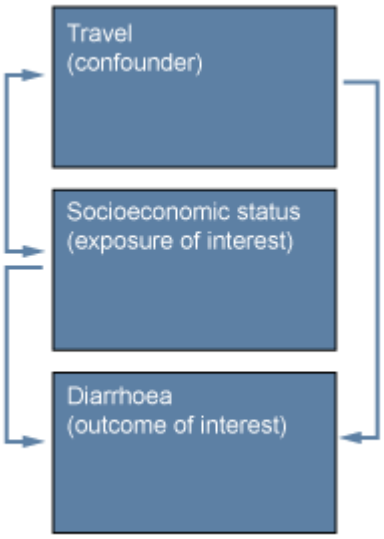
For those that <u>have</u> travelled to the tropics:

Odds ratio = (15*108)/(5*72) = **4.5**

For those that <u>have not</u> travelled to the tropics:

Odds ratio= (144*300)/(336*20) = **6.4**

So when we look at the two groups (travelled to tropics or not travelled to tropics) separately, we see that the association between diarrhoea and socioeconomic group is stronger than when we analysed the all the data together. This is because travel to a tropical country is a risk factor for diarrhoea, and is associated with socioeconomic group. So travel to a tropical country is acting as a (negative) confounder in the association between socioeconomic group and diarrhoea.



Importance of Identifying Potential Confounders

The practical implication of what we have just seen is that when we design a study to determine whether there is an association between a particular exposure and a disease, we need to think, before we start the study, about potential alternative explanations (confounders) for the association.

Example: If we were planning a study on liver cancer, the list of risk factors would include hepatitis B, hepatitis C, and exposure to aflatoxin, as just a few examples.

If we forget to do this, then we may not be able to consider alternative explanations for any association we find. So we need to collect information on all identifiable risk factors for the disease under study.

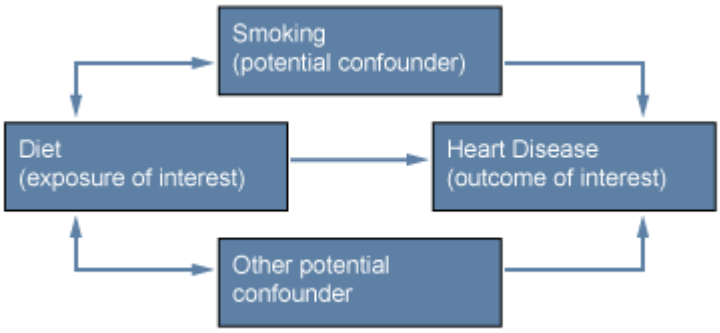
There is no magic way of identifying these risk factors. The main methods are:

1: to think about exposures that are **biologically plausible** as risk factors for the disease in question

2: to do a comprehensive **literature review** to find out what exposures have been found to be risk factors in previous studies.

Any factor which is believed to have an effect on the risk of the disease of interest is a potential confounder.

Smoking is generally believed to be causally related to a wide range of disease (lung cancer, heart disease etc.). So if we were studying the association between any other risk factor (such as diet) and heart disease, we would need to anticipate the possible confounding effect of smoking, and therefore collect data on smoking habits.



Testing for Confounding

There is no specific test that tells us whether a factor has a confounding effect on any given exposure-outcome association. However, if we adjust for a potential confounder in the analysis, and we find that the adjusted estimate of the exposure-outcome association is different from the unadjusted one, this suggests that the crude estimate was confounded by the factor adjusted for.

In our first example of coffee and cancer of the pancreas, the crude (unadjusted) odds ratio was 1.88, but became 1.0 after stratification by smoking.

Crude Data (before stratification)

	Cases (cancer of the pancreas)	Controls
Coffee	450	200
No Coffee	300	250

Odds ratio = 1.88

Stratified Data

		Cases	Controls
Smokers	*Coffee*	400	100
	No Coffee	200	50
Non-smokers	*Coffee*	50	100
	No Coffee	100	200

Odds Ratio for drinking coffee among smokers = 1.0

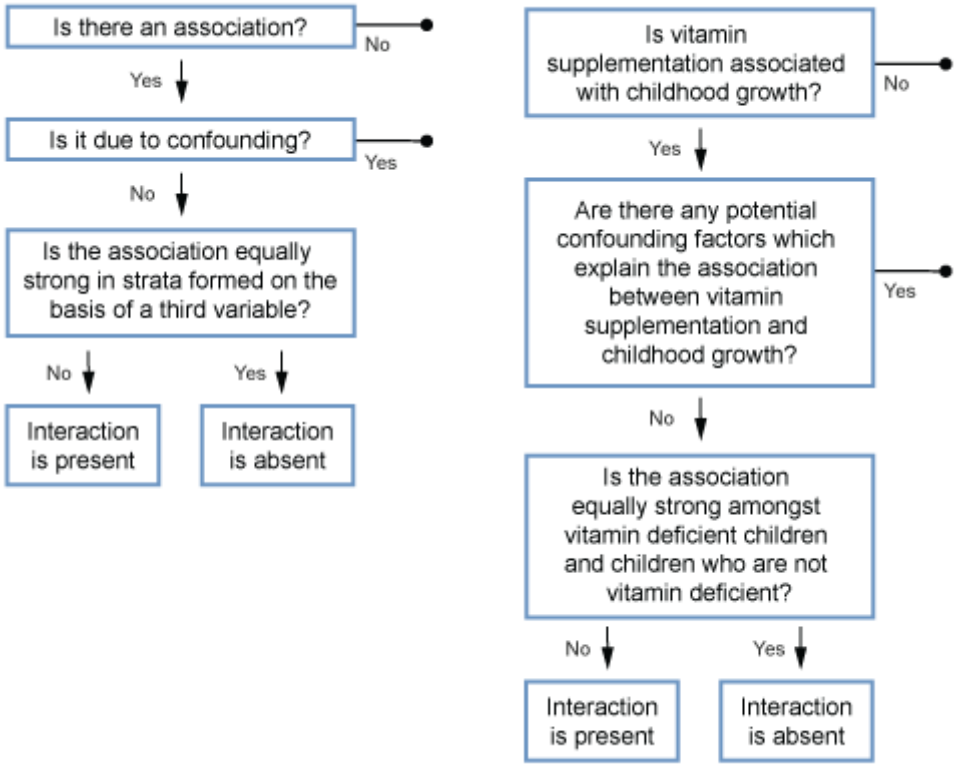
Odds Ratio for drinking coffee among non-smokers = 1.0

Effect Modification vs. Confounding

Suppose we do a study to investigate the effect of vitamin X supplements on childhood growth. Among children who are vitamin X deficient, it is likely that vitamin X supplements will be associated with increased growth. However, among children who are not deficient in vitamin X, supplements may have no effect.

This is an example of **effect modification**, also known as **interaction**. This can be defined as the situation where the association between an exposure (here, vitamin X supplements) and an outcome (childhood growth) varies by levels of a third factor (level of vitamin X before supplementation). You will also see this phenomenon called **effect measure modification** in the literature because its detection depends on the choice of effect measure.

The diagram below shows how we can evaluate the presence of interaction.



Effect modification is not the same as confounding.

Confounding is a nuisance effect, which distorts the association between an exposure of interest and an outcome, because the confounder affects the outcome and is unequally distributed between the subjects who are exposed or unexposed to the exposure of interest. So we aim to control confounding.

Effect modification is a real and interesting effect such that the association between an exposure and an outcome differs according to the level of a third factor. We want to detect and report effect modification, not control it. Exploring the nature of effect modification can be very helpful in understanding the biological processes underlying an association between an exposure and an outcome.

Example: Diabetes as a Risk for Coronary Heart Disease

	Coronary Heart Disease (CHD)	No CHD
Diabetes	25	170
No diabetes	95	2194

Crude odds ratio = (25*2194)/(170*95) = 3.40.

Stratified Data

		CHD	No CHD
Females	*Diabetes*	13	93
	No diabetes	25	1191
Males	*Diabetes*	12	77
	No diabetes	70	1003

Odds Ratio for CHD among women = 6.66

Odds Ratio for CHD among male = 2.23

Stratifying by gender, we can calculate different measures. Look at the odds ratios above. The odds ratio for women is 6.66, compared to the crude odds ratio of 3.40. Therefore, women are at much greater risk of diabetes leading to the incident coronary heart disease. For men, the odds ratio is 2.23.

Is diabetes a risk for incident heart disease in men and in women? Yes. Is it the same level of risk? No. For men the OR is 2.23, for women it is 6.66. The overall estimate is closer to a weighted average of the two stratum specific estimates. Gender modifies the effect of diabetes on incident heart disease. We can see that numerically because the crude odds ratio is more representative of a weighted average of the two groups.

What is the most informative estimate of the risk of diabetes for heart disease? 3.40 is not very informative of the true relationship. What is much more informative is to present the stratum specific analysis.

During data analysis, major confounders and effect modifiers can be identified by comparing stratified results to overall results.

Confounding and effect modification are sometimes considered together because the technique of stratification is useful for both. We use stratification to control confounding; we use stratification to detect and describe effect modification.

In any particular situation, a factor can be a confounder, or an effect modifier, or both, or neither.

So, how do we tell the difference between a confounder and an effect modifier?

1. Estimate a crude (unadjusted) estimate between exposure and disease.
2. Stratify the analysis by any potential major confounders to produce stratum-specific estimates.
3. Compare the crude estimator with stratum-specific estimates and examine the kind of relationships exhibited.

- **With a Confounder:**
  - the crude (unadjusted) estimator (e.g. RR, OR) is outside the range of the two stratum-specific estimators (in the Coffee drinking example - the crude odds ratio was higher than both of the stratum specific ratios);
  - If the adjusted estimator is importantly (not necessarily statistically) different (often 10%) from the crude estimator, the “adjusted variable” is a confounder. In other words, if including the potential confounder changes the estimate of the risk by 10% or more, we consider it important and leave it in the model.
  - Statistical methods (Extended Mantel-Haenszel method, multiple regression, multiple logistic regression, proportional hazards) are available to calculate the “adjusted” estimator, accounting for confounders.
  - **Confounding** is concerned with **"alternative" explanations** for the effect seen between the exposure of interest and the outcome.
  - With confounding, the investigator wishes to **remove (or prevent)** the influence of the confounding factor **in order to get nearer to the "truth".** This is done through choice of study design, or by "adjusting" or "controlling" for the confounding factor using statistical methods.

- **With Effect modifiers:**
  - the crude estimator (e.g. RR, OR) is closer to a weighted average of the stratum-specific estimators;
  - the two stratum-specific estimators differ from each other
  - Report separate stratified models or report an interaction term.
  - **Interaction**, on the other hand, occurs when the presence of one factor **modifies** the effect of another (i.e. the **effect** of the exposure of interest **differs** according to **which category** of the **other factor** is being examined).
  - Interaction is an **important property** of the **relationship between two factors**, and their influence on the disease.
  - Rather than try to eliminate this effect, the investigator wants to **detect and describe interaction** in the **greatest possible detail**.

To review, confounders mask a true effect and effect modifiers means that there is a different effect for different groups.

What happens when a factor is both a confounder and an effect modifier?

It may be the case that a factor is both a confounder and an effect modifier. In this situation, we would report stratum-specific measures of effect in order to illustrate the interaction we have detected. In doing this we have eliminated any confounding effect of the variable. For example, if we detected an interaction with sex in a cohort study and reported sex-specific rate ratios, it would no longer be possible for sex to act as a confounder.

We saw in the confounding section that the way to detect the presence of confounding is to stratify the data according to the potential confounding factor. If the stratum-specific rate ratios (or risk ratios, or odds ratios) are all similar to each other, but are different from the overall rate ratio this is evidence that the factor is confounding the relationship between the exposure and the disease in question*. The rate ratios may vary across strata, but this variation is only due to random error*. The next step is to summarize, or *pool*, these stratum-specific rate ratios to obtain a rate ratio *adjusted* for the confounding factor

When interaction is present there is variation between the stratum-specific rate ratios which is not simply due to chance. It is not informative to pool them to obtain a summary measure. In this situation, it is better to report the stratum specific rate ratios separately. There are ways of including terms for interaction in statistical models, which you will learn about in this course.

In summary, the way to distinguish confounding and interaction is to *stratify* the data according to the factor under investigation: If the stratum-specific rate ratios (or odds ratios or risk ratios) differ from the unstratified rate ratio, and if there is little variation between the stratum-specific rate ratios, this is evidence for confounding. If there is variation in the stratum-specific rate ratios (more than is due to chance), this is evidence of interaction.

Each of the situations we have learnt about are summarised in the examples below.

		Crude odds/ rate/ risk ratio	Odds/rate/risk ratio in stratum 1	Odds/ rate/ risk ratio in stratum 2	Adjusted odds/ rate/ risk ratio
Example 1	No confounding No interaction	3.0	3.0	3.0	3.0
Example 2	Confounding No interaction	3.0	2.0	2.0	2.0
Example 3	Interaction	3.0	0.8	5.5	Should not be calculated

SUMMARY

What is confounding?

Confounding is about alternative explanations for an association between an exposure and an outcome.

Confounding occurs when there is unequal distribution of a risk factor for a disease between those exposed and unexposed to the exposure of interest.

What are potential confounders?

To be a potential confounder, a variable must be:

- associated with the exposure of interest in the source population
- a risk factor for the outcome among the unexposed
- not on the causal pathway between the exposure and outcome of interest

Positive and negative confounding

Confounding can be

- positive i.e. the finding of an apparent association between an exposure and an outcome when there is no true association (or making an association appear to be stronger than it really is)
- negative i.e. the apparent absence of an association when one truly exists (or making an association appear to be weaker than it really is)

Controlling confounding

It is essential to identify potential confounders at the stage of study design: this is done on the basis of biological plausibility and by reviewing the literature.

Methods for controlling confounding at the stage of study design are:

- restriction
- matching
- randomisation

Methods for controlling for confounding at the analysis stage are:

- stratification
- regression modelling

Effect Modification

Effect modification (interaction) describes the situation where the association between an exposure and an outcome varies according to the level of a third factor.

Confounding is a nuisance which we aim to control: effect modification may help us understand an association and we should aim to detect and report it. Whenever effect modification is present, stratified measures of effect are reported.