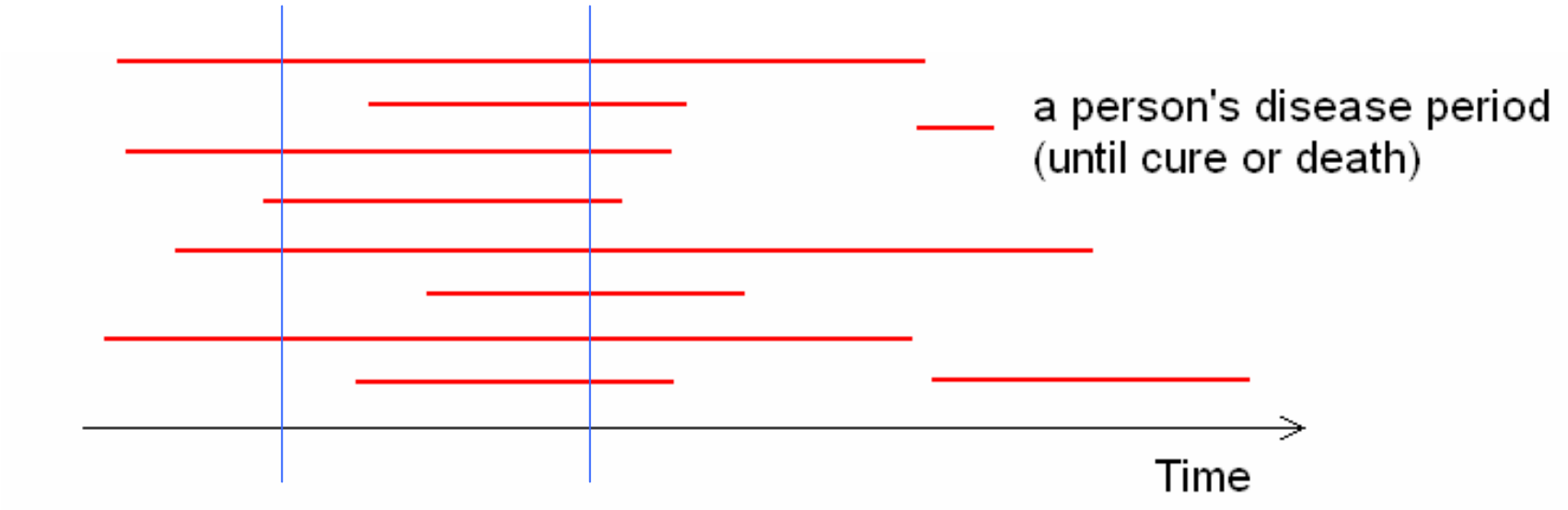
Incidence and prevalence

Occurrence of disease (or condition of interest) has its dynamics in the population.



Cross-sectional, longitudinal, and register-based studies reflect to quite different aspects of this dynamics.

In a **cross-sectional** study, you can calculate the **proportion of those with the disease** at a particular time point.

Prevalence: proportion (or %) of persons with the disease

- in the study population
- at a particular time point

In a **cross-sectional** study, you can calculate the **proportion of those with the disease** at a particular time point.

Prevalence: proportion (or %) of persons with the disease

- in the study population
- at a particular time point

„In Hungary, 55% of the old people suffer from high blood pressure.”

In a **cross-sectional** study, you can calculate the **proportion of those with the disease** at a particular time point.

Prevalence: proportion (or %) of persons with the disease

- in the study population
- at a particular time point

„In Hungary, 55% of the old people suffer from high blood pressure."

What should be added?

- When? (Give the time point of the study)
- How is „old" meant? (Specify the study population)
- What is regarded as high blood pressure? (Specify exactly the condition of interest)

„Prevalence of tuberculosis in Nicaragua in 2007 was 74 cases per 100,000 people.”

(reads better than 0.00074 or 0.074%)

„Prevalence of tuberculosis in Nicaragua in 2007 was 74 cases per 100,000 people.”

(reads better than 0.00074 or 0.074%)

When estimating prevalence from a screening survey, also the **sample size** should be given.

„According to our screening (*n*=2430), the prevalence of tuberculosis is estimated as 120 in 100,000.”

„Prevalence of tuberculosis in Nicaragua in 2007 was 74 cases per 100,000 people.”

(reads better than 0.00074 or 0.074%)

When estimating prevalence from a screening survey, also the **sample size** should be given.

„According to our screening (*n*=2430), the prevalence of tuberculosis is estimated as 120 in 100,000.”

Looking at the numbers we can see that only 3 people in the sample was found to be positive. This may be too few for a reliable estimate.

„Prevalence of tuberculosis in Nicaragua in 2007 was 74 cases per 100,000 people.”

(reads better than 0.00074 or 0.074%)

When estimating prevalence from a screening survey, also the **sample size** should be given.

„According to our screening (*n*=2430), the prevalence of tuberculosis is estimated as 120 in 100,000.”

Looking at the numbers we can see that only 3 people in the sample was found to be positive. This may be too few for a reliable estimate.

120 per 100,000 is the observed **sample prevalence**. 95% **confidence interval for the population prevalence** is from 20 to 380 per 100,000.

When estimating prevalence of a disease, one should be aware that the diagnostic test may have sensitivity and specificity less than 100%.

there may be false positives and/or fals negatives

When estimating prevalence of a disease, one should be aware that the diagnostic test may have sensitivity and specificity less than 100%.

there may be false positives and/or fals negatives

In this case the experienced sample prevalence is called the **apparent prevalence** while the prevalence corrected for test sensitivity and specificity is called the **true prevalence**.

$$prev_a = prev_t \cdot Se + (1 - prev_t) \cdot (1 - Sp)$$

$$prev_t = \frac{prev_a + Sp - 1}{Se + Sp - 1}$$

If a screening test has *Se*=99% and *Sp*=75%, then 40% apparent prevalence results in just 20.2% true prevalence.

In a register-based study, one can count the **new cases in a particular time period**.

Incidence (also called **incidence rate** or **cumulative incidence**) is the number of new cases

- per time period (year, month, etc.)
- per 100,000 people **at risk (in the study population)**

„Incidence of breast cancer in the UK in 2004 among females aged 40-44 was 20 cases per month per 100,000."

Relation between incidence and prevalence

How is it possible that a disease has low incidence yet high prevalence?

Relation between incidence and prevalence

How is it possible that a disease has low incidence yet high prevalence?

this may occur with chronic diseases

Relation between incidence and prevalence

How is it possible that a disease has low incidence yet high prevalence?

this may occur with chronic diseases

How is it possible that a disease has high incidence yet low prevalence?

this occurs with acute diseases (short duration until cure or death)

Relation between incidence and prevalence

How is it possible that a disease has low incidence yet high prevalence?

this may occur with chronic diseases

How is it possible that a disease has high incidence yet low prevalence?

this occurs with acute diseases (short duration until cure or death)

(mean) **prevalence** = (mean) **incidence** · (mean) **duration**

Association, prediction, and causality

Association between two conditions or variables is **any kind of relation** between them, **not necessarily causal**.

Association means that occurrence of a condition *A* is more likely (or less likely) together with *B* than without *B*.

Association, prediction, and causality

Association between two conditions or variables is **any kind of relation** between them, **not necessarily causal**.

Association means that occurrence of a condition *A* is more likely (or less likely) together with *B* than without *B*.

- *A* may be the cause of *B* (directly or indirectly)
- *B* may be the cause of *A* (directly or indirectly)
- A third factor may be the cause of both *A* and *B*

Association, prediction, and causality

Association between two conditions or variables is **any kind of relation** between them, **not necessarily causal**.

Association means that occurrence of a condition *A* is more likely (or less likely) together with *B* than without *B*.

- *A* may be the cause of *B* (directly or indirectly)
- *B* may be the cause of *A* (directly or indirectly)
- A third factor may be the cause of both *A* and *B*

Some spurious correlations: there is a positive correlation between

- ice cream consumption and incidence of drowning deaths
- number of storks and fertility

Association, prediction, and causality

Association between two conditions or variables is **any kind of relation** between them, **not necessarily causal**.

Association means that occurrence of a condition *A* is more likely (or less likely) together with *B* than without *B*.

- *A* may be the cause of *B* (directly or indirectly)
- *B* may be the cause of *A* (directly or indirectly)
- A third factor may be the cause of both *A* and *B*

Some spurious correlations: there is a positive correlation between

- ice cream consumption and incidence of drowning deaths
- number of storks and fertility

What could be the third factor?

Confounding

We are interested in exploring the relationship (some kind of association) between the variables *X* and *Y*.

Confounding is experienced if a third variable *Z* modifies the relationship so that

- it generates an apparent association between *X* and *Y*
- it masks (hides) an existing association between *X* and *Y*
- it modifies the strength and/or direction of the association

In such cases *Z* is called a **confounder**.

Confounding

We are interested in exploring the relationship (some kind of association) between the variables *X* and *Y*.

Confounding is experienced if a third variable *Z* modifies the relationship so that

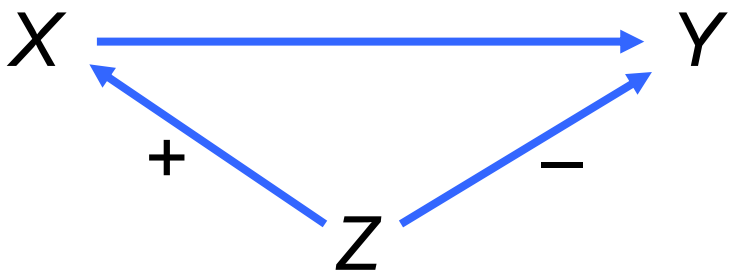
- it generates an apparent association between *X* and *Y*
- it masks (hides) an existing association between *X* and *Y*
- it modifies the strength and/or direction of the association

In such cases *Z* is called a **confounder**.

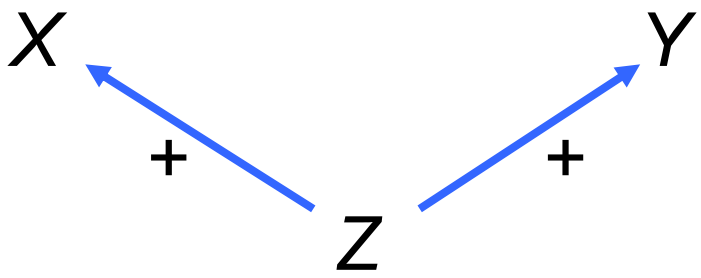
Of course we want to eliminate the confounding effect of *Z* to see the pure (direct) relationship between *X* and *Y*.

But there may be further confounders as well...

Any *Z* can be a confounder only if it is associated (or correlated) with both *X* and *Y*.

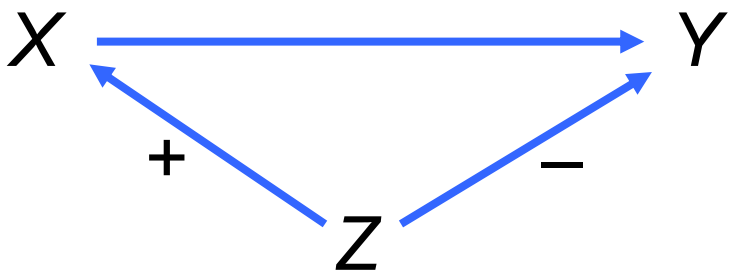


masked existing correl.

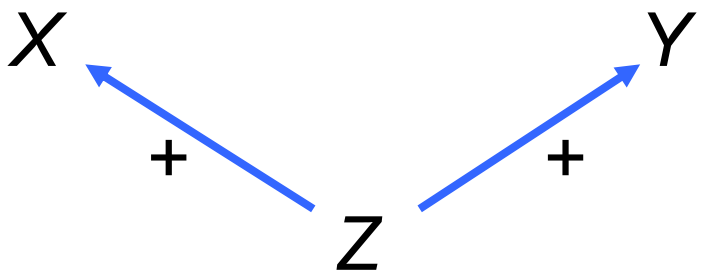


apparent positive correl.

Any *Z* can be a confounder only if it is associated (or correlated) with both *X* and *Y*.



masked existing correl.

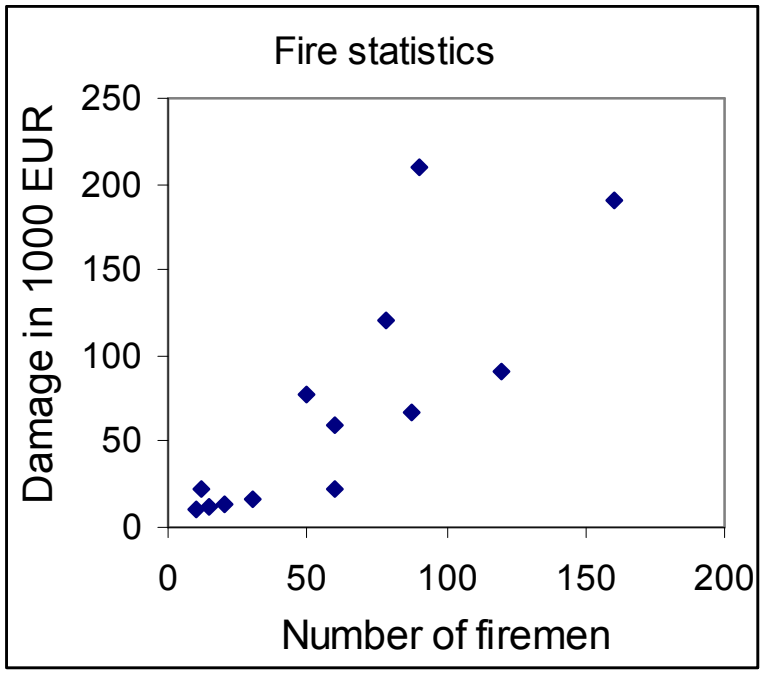


apparent positive correl.

Example:

Strong positive correlation can be observed between the number of firemen and the damages of fire.

Isn't it perhaps worth at all calling the fire brigade?



Prevalence rates, mortality rates etc. can differ between two populations (regions or countries) purely due to the different age structures in the two populations.

	Popul. A		Popul. B	
	Preval.	Age structure	Preval.	Age structure
Young (<30)	5%	20%	5%	40%
Medium (30-60)	15%	40%	20%	40%
Old (>60)	35%	40%	40%	20%
Total	21%		18%	

In each age group

$$\textit{prevalence in Pop. A} \leq \textit{prevalence in Pop. B}$$

holds, still in the whole population the relation is the opposite.

Prevalence rates, mortality rates etc. can differ between two populations (regions or countries) purely due to the different age structures in the two populations.

	Popul. A		Popul. B	
	Preval.	Age structure	Preval.	Age structure
Young (<30)	5%	20%	5%	40%
Medium (30-60)	15%	40%	20%	40%
Old (>60)	35%	40%	40%	20%
Total	21%		18%	

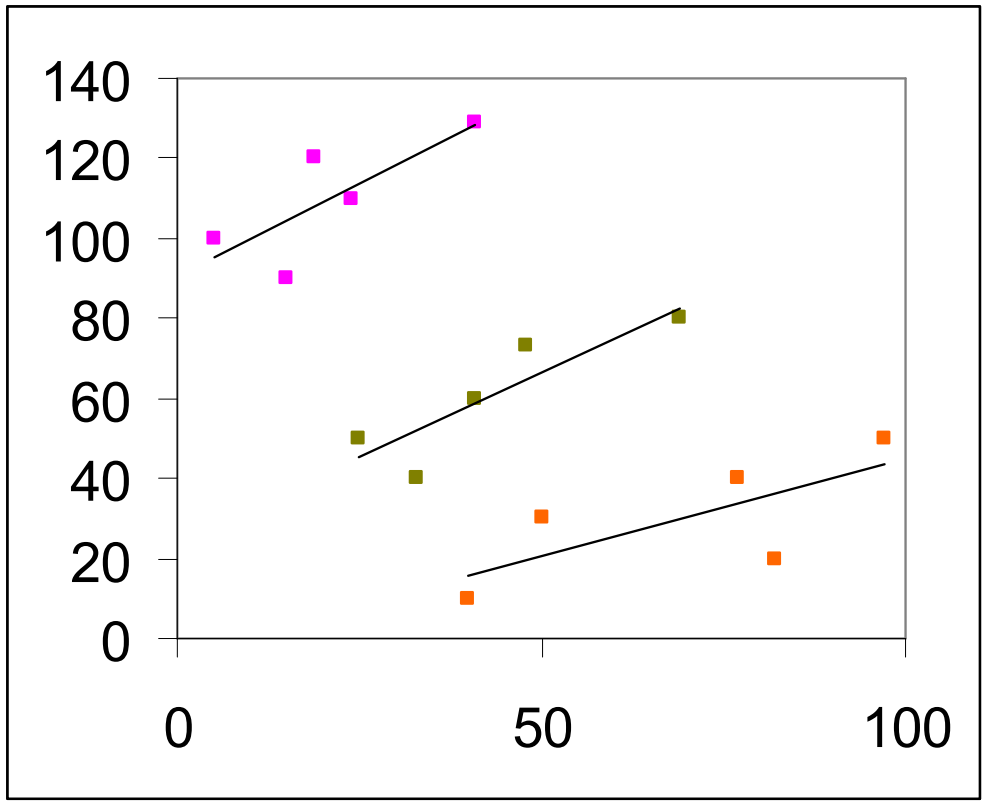
In each age group

$$\textit{prevalence in Pop. A} \leq \textit{prevalence in Pop. B}$$

holds, still in the whole population the relation is the opposite.

Standardization of rates helps avoid such problems!

It may occur that there is a positive correlation between *X* and *Y* in each of some groups, but the pooled sample exhibits a negative correlation. Then the grouping variable is a confounder and should be included in the model.



In health studies age and gender should always be considered as potential confounders.

Further variables that can be taken into account as confounders: severity of disease, center (in multicenter studies), time (children are growing, environment is changing, etc.).

In health studies age and gender should always be considered as potential confounders.

Further variables that can be taken into account as confounders: severity of disease, center (in multicenter studies), time (children are growing, environment is changing, etc.).

If you observe an association or correlation that is nonsense, look for a potential confounder!

Also, if an association that should be there is missing!

Methods to eliminate (or at least reduce) confounding

- Matching (for each case we assign a control person of same age, sex, etc.)
- Narrowing the range of inclusion criteria (we include only males aged 40 to 45 in the study)
- Stratifying by the potential confounders and either analysing within each stratum, or using combination methods (standardization of rates, Mantel-Haenszel, etc.)
- Including the potential confounders as covariates in regression-type models (ANOVA, ANCOVA, logistic regression, etc.)

Methods to eliminate (or at least reduce) confounding

- Matching (for each case we assign a control person of same age, sex, etc.)
- Narrowing the range of inclusion criteria (we include only males aged 40 to 45 in the study)
- Stratifying by the potential confounders and either analysing within each stratum, or using combination methods (standardization of rates, Mantel-Haenszel, etc.)
- Including the potential confounders as covariates in regression-type models (ANOVA, ANCOVA, logistic regression, etc.)

And what happens to the unknown (unmeasured) confounders?

Sources of bias

Bias is a discrepancy between a finding and the truth.

A **measurement method is biased** if the measured value (systematically) differs from the true value.

Sources of bias

Bias is a discrepancy between a finding and the truth.

A **measurement method is biased** if the measured value (systematically) differs from the true value.

To detect the bias, we have to know the true value!!!

Sources of bias

Bias is a discrepancy between a finding and the truth.

A **measurement method is biased** if the measured value (systematically) differs from the true value.

To detect the bias, we have to know the true value!!!

Uncontrolled confounders can cause a bias: e.g. change a positive correlation to negative. By controlling for the confounders, we can eliminate the bias in the correlation coefficient.

Sources of bias

Bias is a discrepancy between a finding and the truth.

A **measurement method is biased** if the measured value (systematically) differs from the true value.

To detect the bias, we have to know the true value!!!

Uncontrolled confounders can cause a bias: e.g. change a positive correlation to negative. By controlling for the confounders, we can eliminate the bias in the correlation coefficient.

An **estimation method is biased** if it (systematically) differs form the true parameter. Ignoring *Se* and/or *Sp* of the diagnostic test, prevalence estimate may be biased.

A **sample is biased** if it does not mirror the target population (i.e. if it is not representative). Selection bias is a feature of the sampling process resulting in a biased sample.

A **sample is biased** if it does not mirror the target population (i.e. if it is not representative). Selection bias is a feature of the sampling process resulting in a biased sample.

"…Controls were selected randomly from those visited the same hostpital and underwent a laboratory examination in 2002."

Does this sample mirror well the healthy population?

A **sample is biased** if it does not mirror the target population (i.e. if it is not representative). Selection bias is a feature of the sampling process resulting in a biased sample.

"…Controls were selected randomly from those visited the same hostpital and underwent a laboratory examination in 2002."

Does this sample mirror well the healthy population?

Published results on a topic tend to be biased toward significant results. This is called publication bias.

A **sample is biased** if it does not mirror the target population (i.e. if it is not representative). Selection bias is a feature of the sampling process resulting in a biased sample.

"…Controls were selected randomly from those visited the same hostpital and underwent a laboratory examination in 2002."

Does this sample mirror well the healthy population?

Published results on a topic tend to be biased toward significant results. This is called publication bias.

The word bias is not a well-defined statistical term. It is used in several different meanings. These were just examples from the spectrum.

Interaction

Interaction qualifies the joint effect of two (or more) factors to a dependent variable.

In the simplest case, the joint effect is simply the sum of the two effects: this is called **additivity** of effects, or **no interaction**. Any other joint effect is called **interaction**.

Interaction

Interaction qualifies the joint effect of two (or more) factors to a dependent variable.

In the simplest case, the joint effect is simply the sum of the two effects: this is called **additivity** of effects, or **no interaction**. Any other joint effect is called **interaction**.

So what you should memorize is the definition of no interaction!

Interaction

Interaction qualifies the joint effect of two (or more) factors to a dependent variable.

In the simplest case, the joint effect is simply the sum of the two effects: this is called **additivity** of effects, or **no interaction**. Any other joint effect is called **interaction**.

So what you should memorize is the definition of no interaction!

Assume that treatment *A* causes a decrease in blood glucose level by 3 mmol/l while treatment B causes an increase by 2 mmol/l.

If there is no interaction between the treatments, their joint application will reduce the blood glucose level by 1 mmol/l.

Another (equivalent) formulation of “no interaction”:

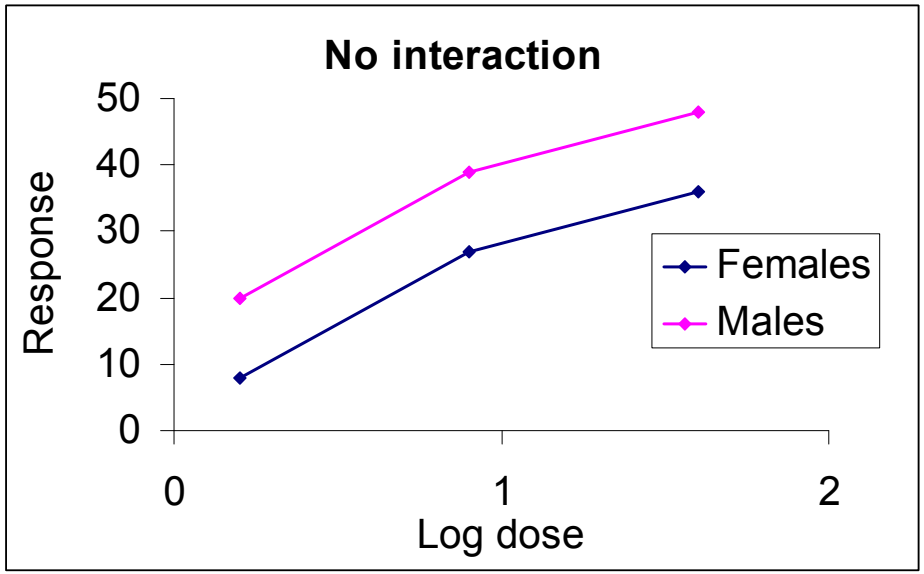
There is **no interaction** between the effects of treatments *A* and *B* if **the effect of *B* is same in each treatment category of *A***.

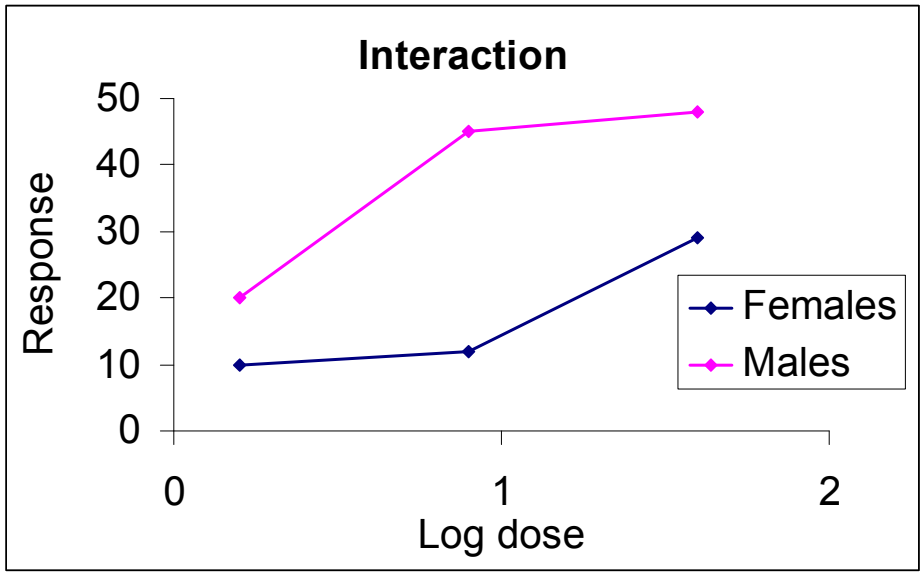
Another (equivalent) formulation of “no interaction”:

There is **no interaction** between the effects of treatments *A* and *B* if **the effect of *B* is same in each treatment category of *A***.

Assume *A* represents 3 doses of a drug, and *B* is sex. No interaction means, that the effect of sex, i.e. the difference in response between females and males is same for each dose.

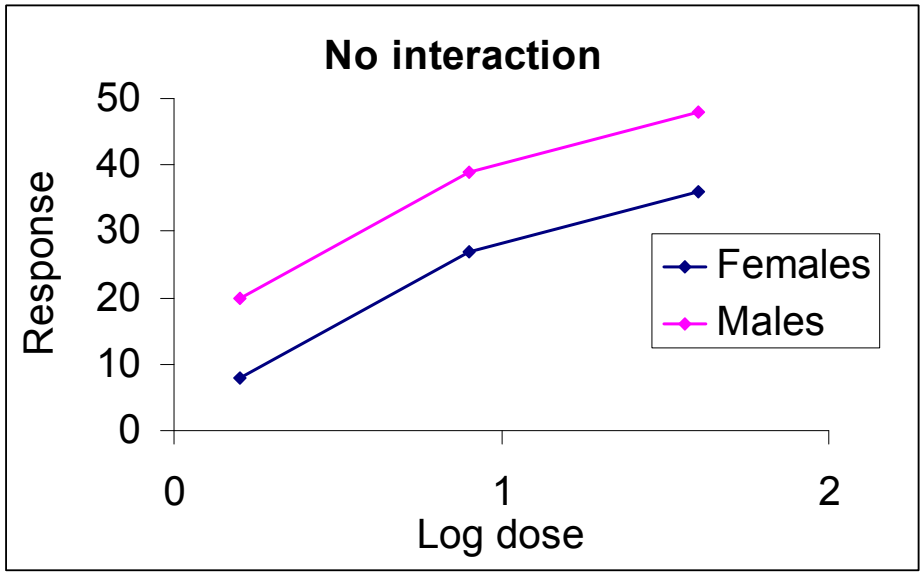
Interaction plot serves as a nice tool to illustrate interactions.

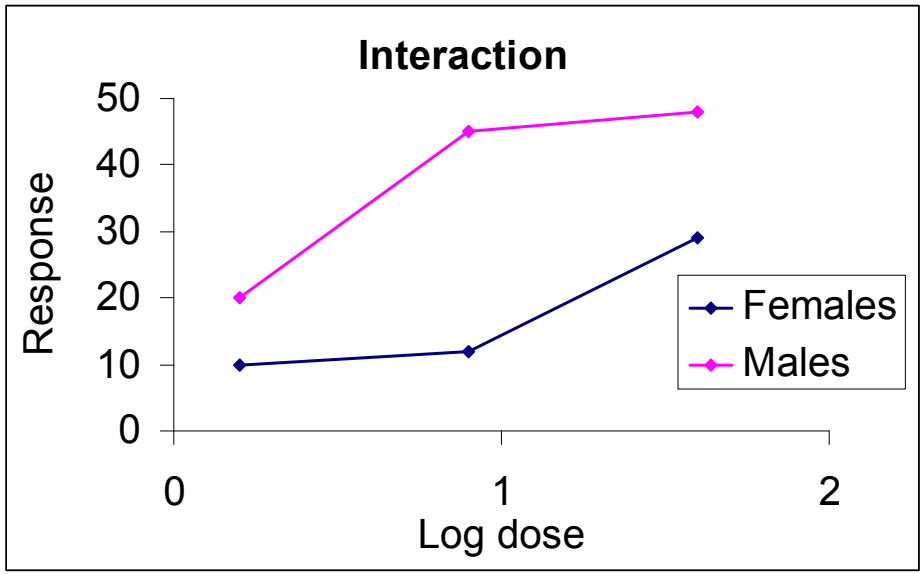




In case of an interaction between dose and sex, an analysis ignoring sex results in an “average” dose-response curve. (Then sex will act then as an uncontrolled confounder.)

Interaction plot serves as a nice tool to illustrate interactions.





In case of an interaction between dose and sex, an analysis ignoring sex results in an "average" dose-response curve. (Then sex will act then as an uncontrolled confounder.)

Don't think that presence of interaction is needed for confounding!

Sex can be a confounder even if there is no interaction!