Bias and Confounding

Dr. Sireen Alkhaldi

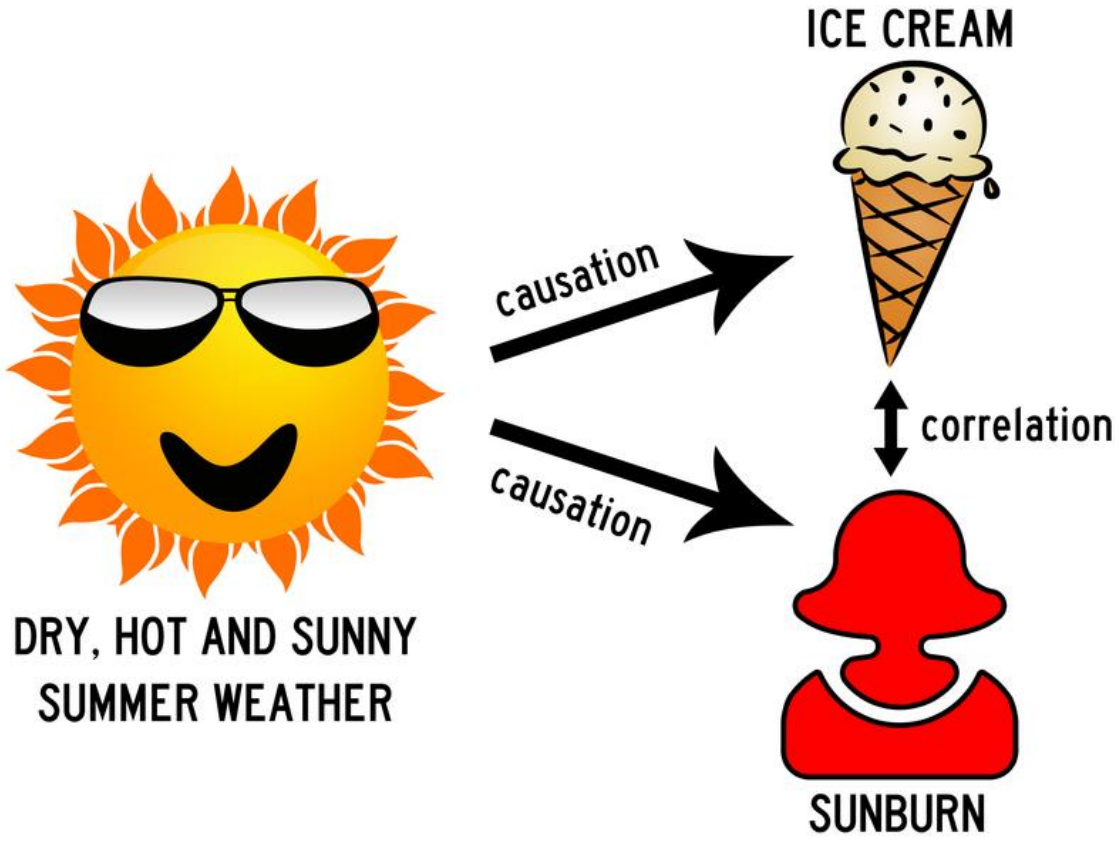
Epidemiology and Biostatistics 2018/2019

School of Medicine/ The University of Jordan

If an association is observed, the first question asked must always be …

“Is it real?”

Remember, correlation is not causation



Ice-cream does not cause sunburn!!!

People often say correlation when they mean association

Bias and Confounding in Epidemiological studies

While the results of an epidemiological study may reflect the true effect of an exposure(s) on the development of the outcome under investigation, the findings may in fact be due to an alternative explanation.

Bias and Confounding in Epidemiological studies

Such alternative explanations may be due to the effects of **bias or confounding** which may produce spurious results, leading us to conclude:

a. The existence of a valid statistical association when one does not exist

b. The absence of an association when one is truly present.

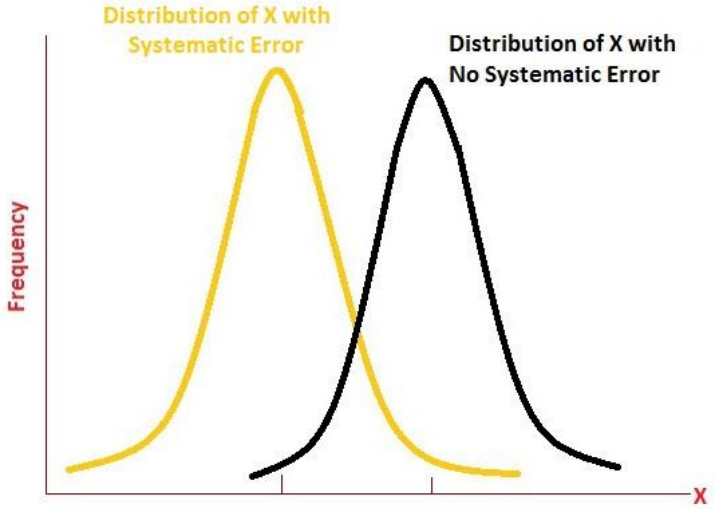
The effect of bias and confounding cannot be eliminated entirely, but can be minimized if the researcher is aware of the potential types of biases and confounders that might distort study results.

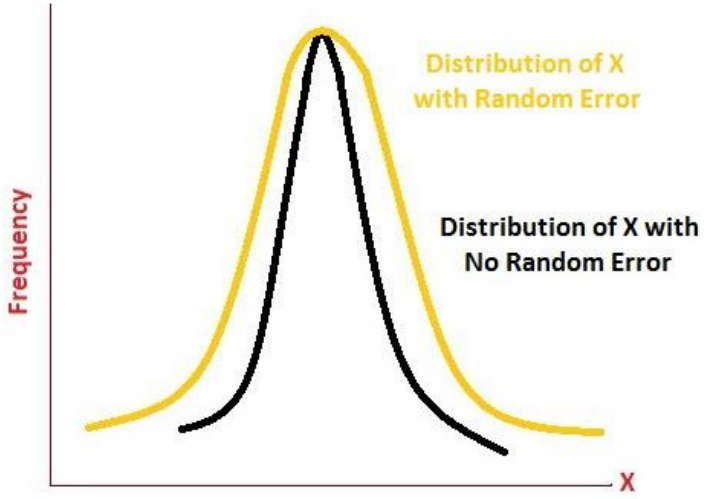
Bias

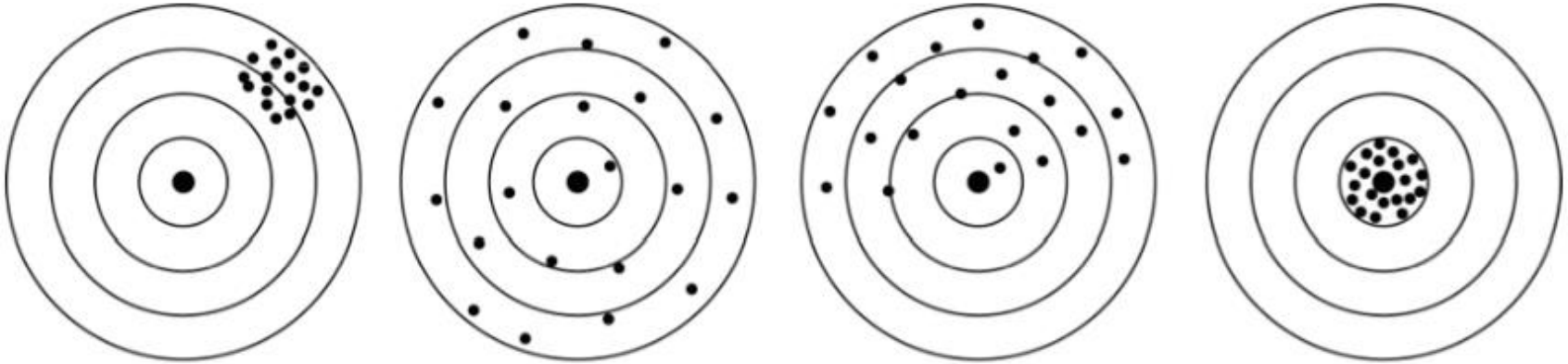
Bias is a systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease (Schlesselman and Stolley, 1982).

"Error" indicates that it is most probably unintentional.
"Systematic " implies that once it is introduced into the study, it cannot be fixed or removed.

✓ The effect of bias will be an estimate either above or below the true value, depending on the direction of the systematic error. So, it affects the **validity** of the study (the degree to which the measurement reflects the true value in the population).







reliable
not valid
(precise,
inaccurate data set)

valid
not reliable
(accurate,
imprecise data set)

not reliable
not valid
(inaccurate,
imprecise data set)

reliable
valid
(accurate,
precise data set)

Accuracy: closeness to the true value (validity) / Precision: closeness to each other (reliability).

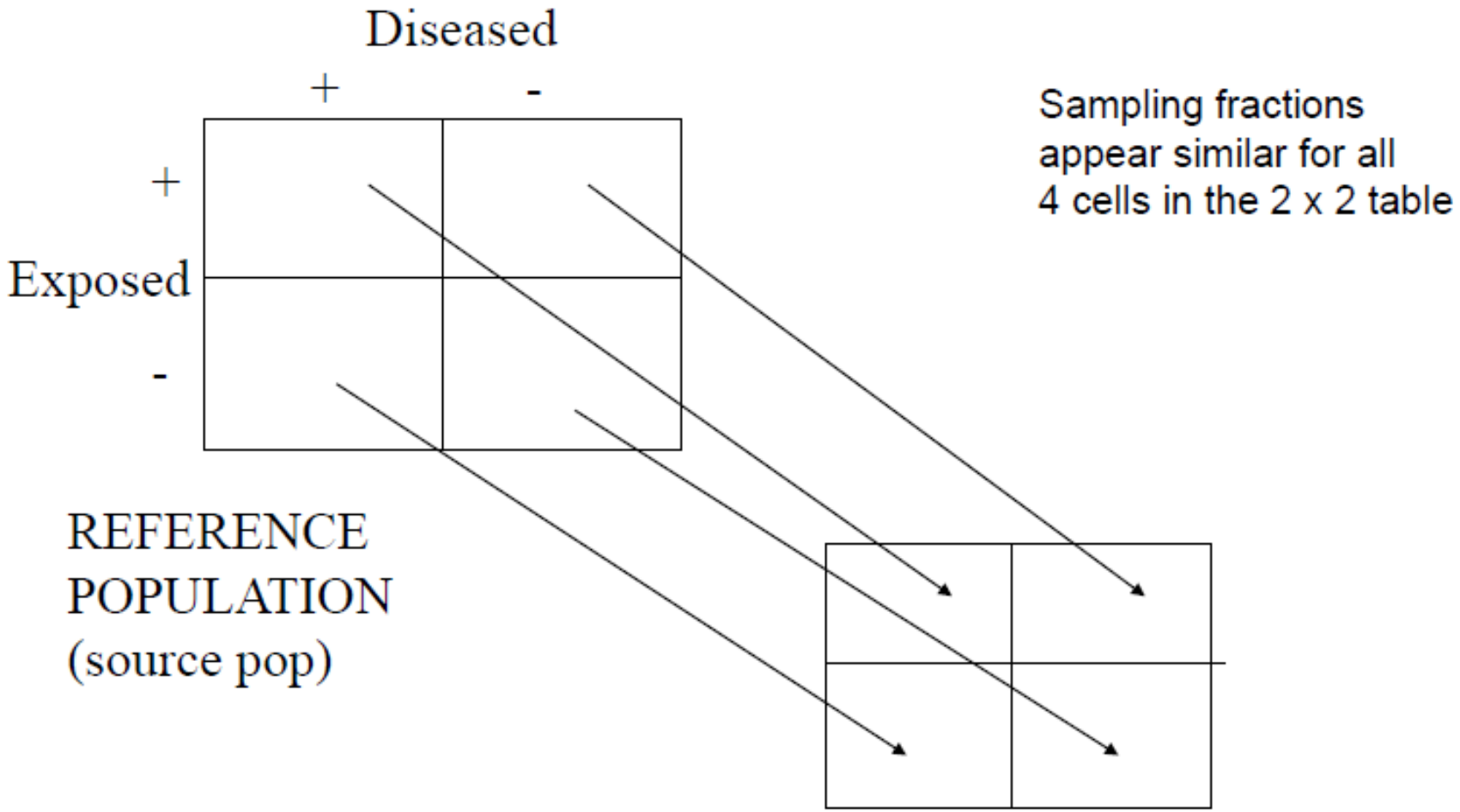
Types of Bias

Common types of bias in epidemiological studies

More than 50 types of bias have been identified in epidemiological studies, but for simplicity they can be broadly grouped into two categories:

1. **Information bias and**
2. **Selection bias**

Unbiased Sampling



Selection Bias

Selection bias is a method of participant selection that distorts the exposure-outcome relationship from that present in the target population

Selection bias occurs when there is a systematic difference between either:

- Those who participate in the study and those who do not (affecting generalisability) *or*
- Those in the treatment group of a study and those in the control group (affecting comparability between groups).

Sources of Selection Bias

Sampling bias describes the scenario in which some individuals within a target population are more likely to be selected for inclusion than others

Volunteer bias: (self-selection bias).

Selecting volunteers to participate in a study:

1. Volunteers could be more health conscious than non-volunteers, thus resulting in less disease

2. Volunteers could also be at higher risk, such as having a family history of illness, thus resulting in more disease

Sources of Selection Bias

Loss to follow-up is a particular problem associated with cohort studies. Bias may be introduced if the individuals lost to follow-up differ with respect to the exposure and outcome from those persons who remain in the study. The differential loss of participants from groups of a randomized control trial is known as attrition bias.

The healthy worker effect is a potential form of selection bias specific to occupational cohort studies. For example, an occupational cohort study might seek to compare disease rates amongst individuals from a particular occupational group with individuals in an external standard population. There is a risk of bias here because individuals who are employed generally have to be healthy in order to work.

Sources of selection bias

Allocation bias occurs in controlled trials when there is a systematic difference between participants in study groups (other than the intervention being studied). This can be avoided by randomization.

Non-response bias: occurs because individuals who do not respond to a call or mailed questionnaire to participate in research studies are generally from those who respond. Example smokers are less likely to return questionnaire than are non-smoker

Exclusion bias: occurs when in certain circumstances epidemiologic studies exclude participants to prevent confounding. Example when exclusion criteria is different for cases and control, or exposed and non-exposed

Sources of Selection Bias

- Selection bias is a particular problem inherent in case-control studies, where it gives rise to non-comparability between cases and controls. In case-control studies, controls should be drawn from the same population as the cases, so they are representative of the population which produced the cases.
- Selection bias can be less of problem in cohort studies compared with case-control studies, because exposed and unexposed individuals are enrolled before they develop the outcome of interest.
- Randomized trials are theoretically less likely to be affected by selection bias, because individuals are randomly allocated to the groups being compared, and steps should be taken to minimize the ability of investigators or participants to influence this allocation process (using randomization).

Minimizing Selection Bias

- Define criteria of selection of diseased and non-diseased participants independent of exposures in a case-control study
- Define criteria of selection of exposed and non-exposed participants independent of disease outcomes in a cohort study
- Use randomized clinical trials

Information Bias

Information bias results from systematic differences in the way data (information) on exposure or outcome are obtained from the various study groups (exposed vs non-exposed) (diseased vs non-diseased).

- ✓ This yields systemic errors in the measurement of exposure or outcome. This will affect the nature of true association.

Information Bias

1. Observer bias:

may be a result of the investigator’s prior knowledge of the hypothesis under investigation or knowledge of an individual's exposure or disease status. Such information may influence the way information is collected, measured or interpretation by the investigator for each of the study groups.

- For example, in a trial of a new medication to treat hypertension, if the investigator is aware which treatment arm participants were allocated to, this may influence their reading of blood pressure measurements. Observers may underestimate the blood pressure in those who have been treated, and overestimate it in those in the control group.

Information Bias

2. Interviewer bias: Occurs where an interviewer asks leading questions that may systematically influence the responses given by interviewees…. Because of the interviewer's knowledge of the participants' exposure or disease status.

Interviewer may probe differently about exposures in the past if he or she knows the subjects as cases.

3. Recall bias: Those with a particular outcome or exposure may remember events more clearly. For example mother of a baby born with congenital disorder remembers every events more clearly during pregnancy than mother with normal baby.

4. Hawthorne effect: people act differently if they know they are being watched.

Remember the Howthorne study??... at a factory to discover if change in lighting would affect productivity. Productivity did increase but only because of increased attention due to the study. As soon as study had ended productivity decreased again

Information Bias

5. Loss to follow-up: those that are lost to follow up or who withdraw from study may be different from those who are followed for the entire study. This can produce biases in the estimate of association

6. Surveillance bias: the group with known exposure or outcome may be followed more closely or longer than the comparison group.

7. Social desirability bias occurs where respondents to surveys tend to answer in a manner they feel will be seen as favorable by others, for example by over-reporting positive behaviours or under-reporting undesirable ones.

In **reporting bias**, individuals may selectively suppress or reveal information, for similar reasons (for example, around smoking history). Reporting bias can also refer to selective outcome reporting by study authors.

Information Bias

8. Performance bias refers to when study personnel or participants modify their behavior / responses where they are aware of group allocations.

9. Detection bias occurs where the way in which outcome information is collected differs between groups.

10. Instrument bias refers to where an inadequately calibrated measuring instrument systematically over/underestimates measurement. The use of standardized, calibrated instruments may reduce the risk of this.

Minimizing Information Bias

1) Observers should be blinded to the exposure and disease status of the individual
2) Blind observers to the hypothesis under investigation.
3) In a randomized controlled trial blind investigators and participants to treatment and control group (double-blinding).
4) Use of a protocol for the collection, measurement and interpretation of information.
5) Use of standardized questionnaires or calibrated instruments, such as sphygmomanometers.
6) Training of interviewers.

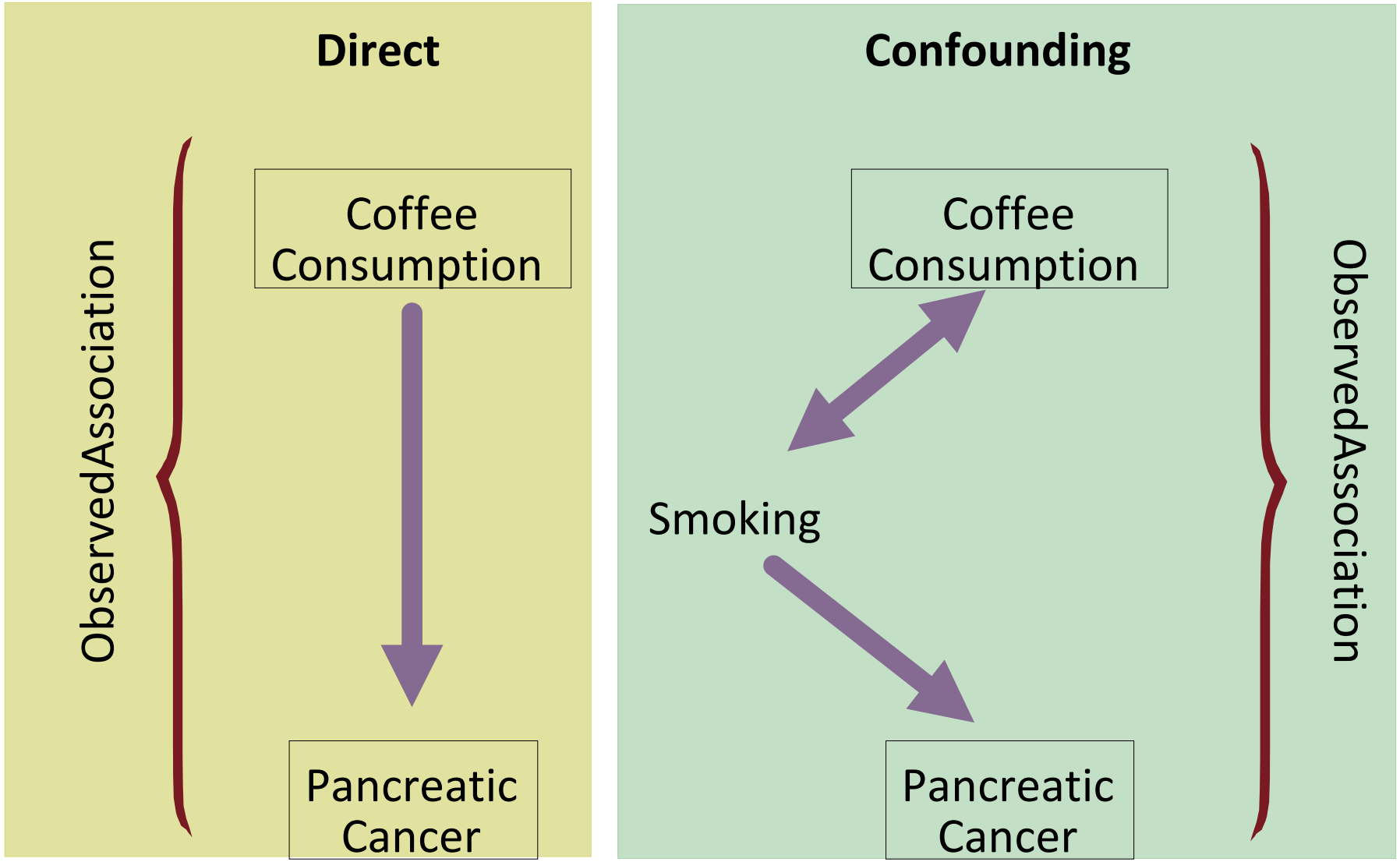
Confounding

Confounding occurs when the observed association between exposure and disease differs from the truth because of the influence of the third variable.

- Effect of a factor of interest is mingled with (confounded with) the effect of another factor.
- Confounder must be:

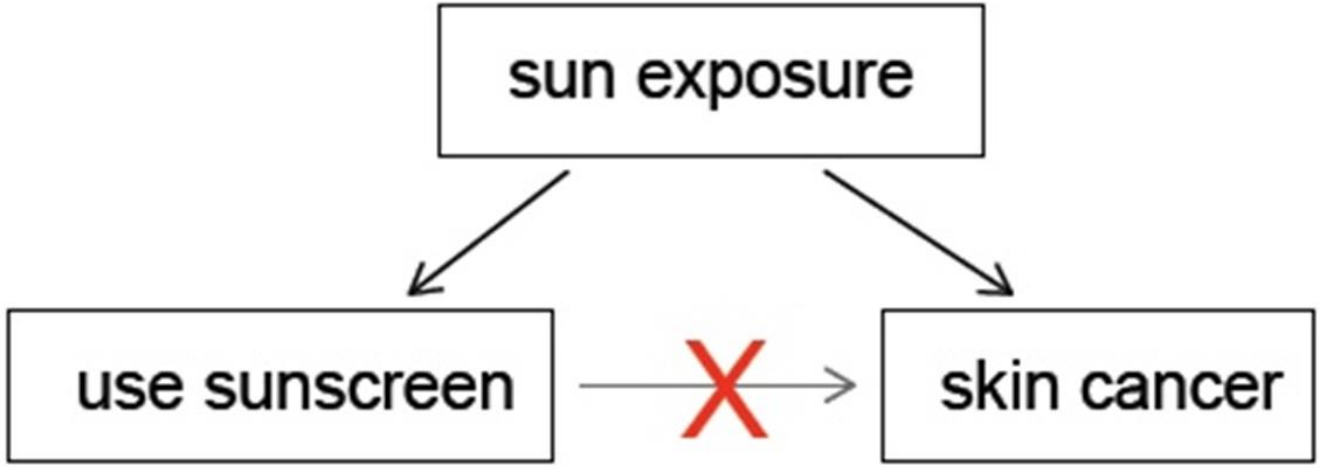
1. Risk factor for the disease independently
2. Associated with exposure under study
3. The variable should not lie on the causal pathway between exposure and disease.

Confounding



Confounding

Confounding may lead to errors in the conclusion of a study, but, when confounding variables are known, the effect may be fixed (removed, corrected, accounted for, controlled for).



Control of Confounding

At design stage:

Restriction: Subject chosen for study are restricted to only those possessing a narrow range of characteristics , to equalize important extraneous factors. e.g. restrict study to women over 50 years.

Matching: for each patient in one group there is one or more patients in comparison group with same characteristics, except for the factor of interest.

Randomization: subjects of study are randomly selected to even out unknown confounders.

Control of Confounding

At analysis stage:

Stratification: Forming strata on the basis of the confounding variable is one method of analysis (age group , socioeconomic status etc).

Multivariate analysis: The statistical analysis of data collected on more than one variable.

Standardization: accounts for confounders (generally age and sex) to counteract the effect of differences in the distribution of confounding factors.