Case-control studies

Dr. Karim Abawi

Geneva Foundation for Medical Education and Research

Training course in research methodology,
research protocol development and scientific writing
Geneva, 2024

Case-control studies

- Case-control studies are observational studies, where two groups determine the level of exposure to a risk or a disease, by identifying a group of individuals with disease and for purpose of comparison, a group of people without the disease.
- The investigator collects retrospective information about exposure to the risk factor from both groups.
- Case-control studies identify subjects by outcome.
- Subjects with disease (or outcome) are called **cases** and subjects without disease (or outcome) are called **controls**.

Design of a case-control study

Study hypothesis

A specific hypothesis is the first and the most important step in an epidemiologic study design.

It must be clearly stated prior to the design of the case–control study.

Poor hypotheses lead to poor study design and problems in interpretation of results.

Definition and selection of cases

- Case definition, based on precise and objective criteria is essential. For example, a histologically confirmed diagnosis should be required for most cancers.
- With precise case definitions, the investigator avoids ambiguity about types of cases and stages of disease to be included or excluded from the study.
- In case-control studies, the investigator should clearly define the eligibility criteria for both cases and controls (inclusion and exclusion criteria).
- It is important to select both cases and controls from the same source of the population. This help to maximize the chance of detecting important association.

Source of cases

- Patient attending health facilities
- Medical records
- Death certificates
- Disease registries (e.g., cancer or birth defect registries)
- Cross-sectional surveys

Definition and selection of controls

Controls must fulfil the same eligibility criteria defined for the cases, with the exception of the disease (outcome). For example, if the cases are women with cervical cancer, over 50 years-old and more, the controls must be selected from women in the same age group without the disease.

Source of controls

Controls should be selected from the population from which the cases are selected. This helps to provide an estimate of the exposure prevalence in the population from which the cases arise.

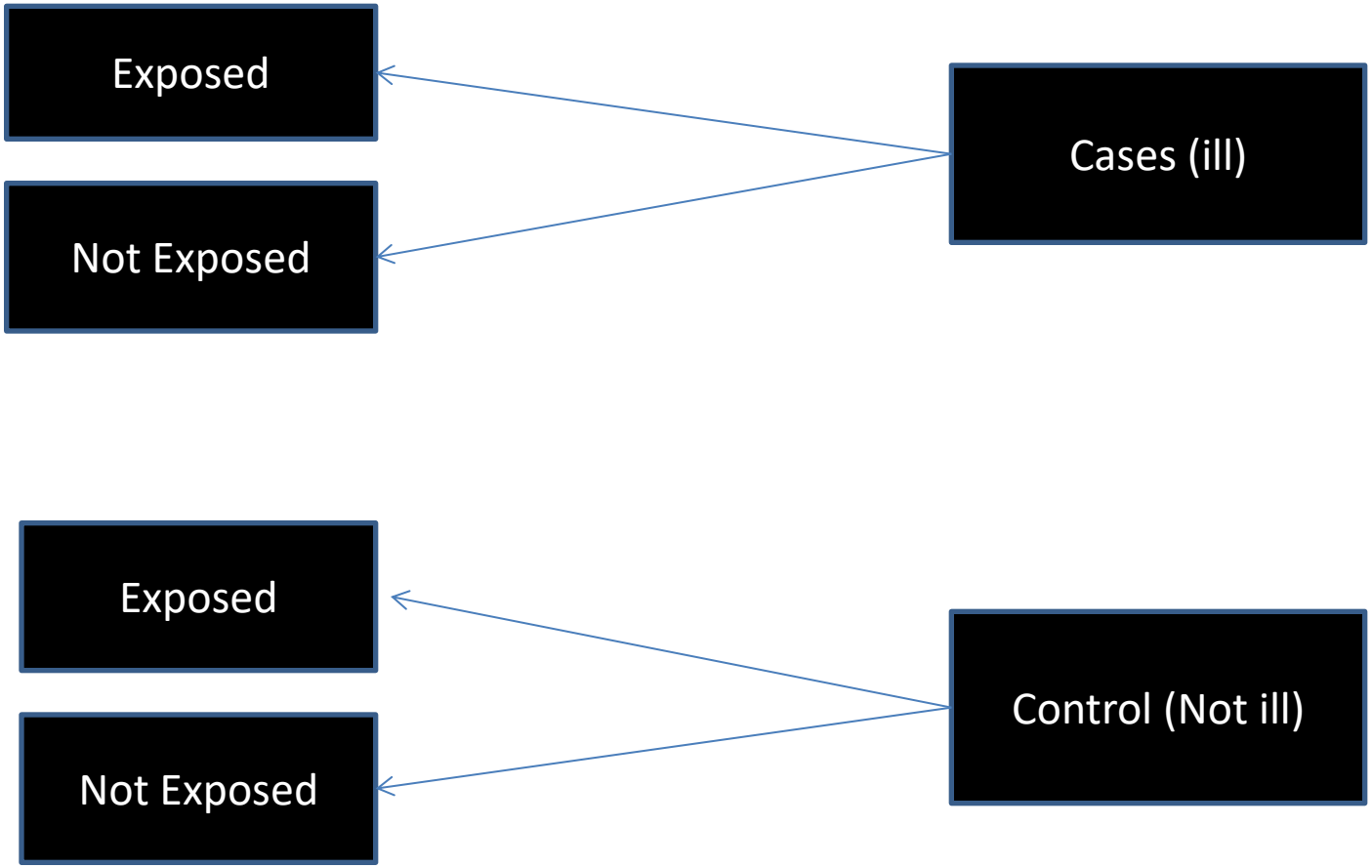
Matching

- It is a method to ensure the comparability between cases and controls.
- Each case is paired individually with a control according to background variables.
- Age, sex, race, sociocultural factors are often used to match cases and controls.
- Appropriate matching helps to reduce the confounders.
- Confounders are variables associated with the both the exposure (risk factor) and the disease (outcome), but is not on the causal pathway between the two.
- Confounders may provide alternative explanations for any associations observed.

Data collection

- Data must be collected in the same way from both groups: cases and controls.
- Investigators must be objective in the search for exposure, especially since the outcome is already known.
- Sometimes it is necessary to interview patients about potential factors, such as smoking history, use of medicine. It may be difficult for some people to recall these details accurately.
- Study participants with the disease (cases) remember details of exposure in the past better than participants without (controls).

Case control study design



Schema of how case-control is conducted:

	Cases (With disease)	Controls (without disease)
Exposed	a	b
Not exposed	c	d
Total	a + c	b + d
Proportions exposed	a /a + c	b/ b +d

Measure of association in case-control studies

In case-control studies the odds ratio is used as a measure of association between exposure and outcome.

Odds ratio is the ratio of odds of exposure in cases to the odds of exposure in controls. In other words, it is a measure of association between an exposure and an outcome. It is calculated as follow:

	Disease		Total
	Yes	No	
Exposed	a	b	a+b
Not exposed	c	d	c+d
Total	a+c	b+d	N

ODDS of EXPOSURE in cases = a/c

ODDS of EXPOSURE in controls = b/d

ODDS RATIO = ratio of odds of exposure = (a/c)/(b/d) = ad/bc

Example: Association between IUD use and PID

	Disease		Total
	Yes	No	
Exposed	841	518	1359
Not exposed	724	967	1691
Total	1565	1485	3050

ODDS RATIO = (a/c)/(b/d) = ad/bc= 841 x 967 /724x 518 =2.2

Interpretation of ODDS ratio

An odds ratio of :

- **1.0 (or close to 1.0):** There is no association between exposure and the disease.
- **Greater than 1.0:** The exposure might be a **risk factor** for the disease.
- **Less than 1.0**: The exposure might have a **protective effect on** the disease.

The magnitude of the odds ratio indicates the strength of association between exposure and the outcome (disease).

Biases in case-control studies

Selection bias

Selection bias occurs when the subjects in one group are different, or the cases and controls are not comparable (other than disease). In order to prevent this bias, precise selection criteria should be defined for both cases and controls.

Ascertainment bias

It may happen because:

•Cases may recall exposure better than the controls

•Investigators may search for exposure better in cases than in control.

Limitations for recalling past events

In case-control studies much data is collected from interviews. Human beings differ in their capacity to recall information. As mentioned above, cases may have better recall than controls. It is also possible that the person may not have the information requested.

Confounding

It occurs when the observed result between exposure and disease is distorted because of the influence of the third variable.

Strengths and weaknesses of case-control studies

Strengths

- Case-control studies cost less than other studies, e.g. cohort studies.
- Case-control studies are mostly retrospective; cases are identified at the beginning. There is no need to follow cases over a period of time, unlike cohort studies.
- Case-control studies are more appropriate for rare diseases.
- The association between diseases and multiple exposures can be studied at the same time.

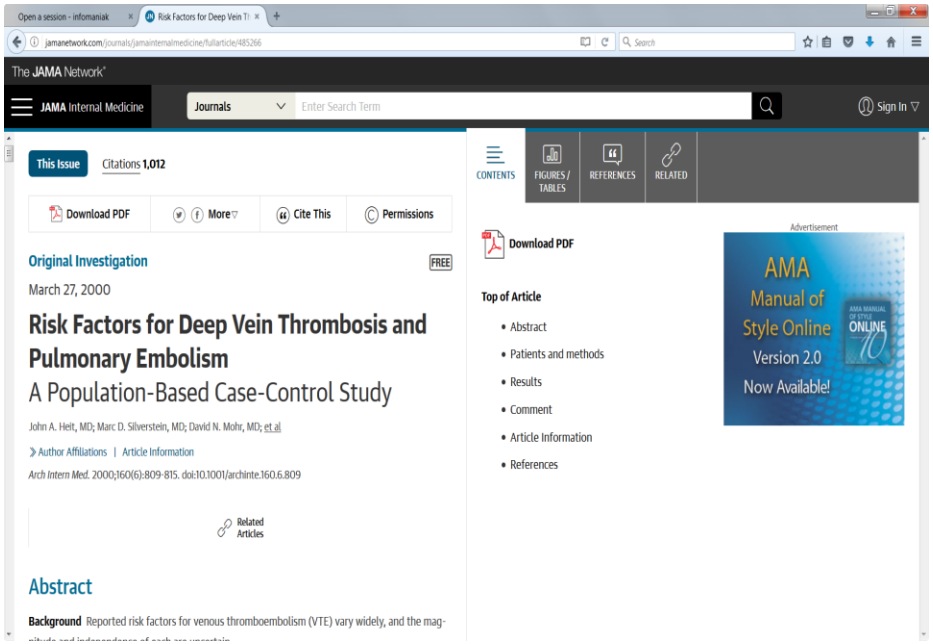
Weaknesses

- Case-controls are subject to multiple biases (especially selection and recall biases).
- Case-control studies cannot estimate the incidence rate unless the rates are population based.
- Case-control studies are difficult for determining the time period between the exposure and disease.

The link below provides access to an example of a case-control study

Article

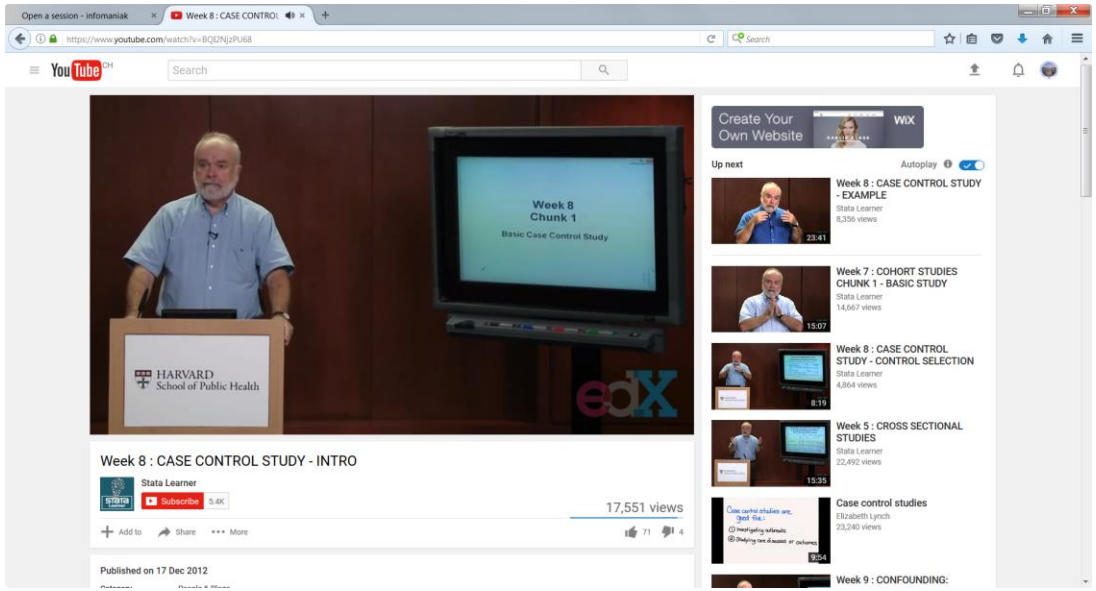
Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ. Risk Factors for Deep Vein Thrombosis and Pulmonary Embolism: A Population-Based Case-Control Study. Arch Intern Med. 2000 Mar 27;160(6):809-15. https://dx.doi.org/10.1001/archinte.160.6.809



The link below provides access to a teaching session on case-control studies

Video

https://www.youtube.com/watch?v=BQI2NjzPU68



References

- Aschengrau A, Seage GR. Essentials of epidemiology in public health. Jones & Bartlett Publishers; 2008.
- CDC. Reproductive Health- Data and Statistics: Epidemiology Glossary [Internet]. CDC. 2015. https://www.cdc.gov/reproductive-health/glossary/index.html
- Evidence Based Medicine Toolkit: Clinical Epidemiology Glossary [Internet]. Buckingham Jeanette, Fisher Bruce, Saunders Duncan. c2008. http://www.ebm.med.ualberta.ca/Glossary.html
- Gordis L. Epidemiology. 4th ed. Saunders; 2008.