



One Minute Revision Guides

**Study design in epidemiology**

Author: Dr Reuben Arasaratnam

# Study design in epidemiology

## Objectives

- Understand the principles of bias, chance and confounding in study designs.
- Be able to list study designs and their strengths and weaknesses
- Be able to calculate the appropriate measures of association for each study

## The aim of a good study

The majority of the decisions that we make in medical practice are based on evidence from population studies. Information from studies can provide us with:

- risk factors for a disease
- comparative efficacy of treatments
- adversity profile of drugs

When a study claims to have to show a particular effect, how do we know it's true? For example, if I claimed that one study showed that drinking alcohol increased the risk of lung cancer how could you assess the validity of this claim?

There are a number of flaws in study design that could mean the claim from my study **is not true**. Here are a few important considerations:

### 1. Could the result have happened by chance?

Is it just luck? To answer this question you should look at the statistics- especially the p values and the confidence intervals of the results. A statistically significant result is usually indicated by the statement  $p < 0.05$  or by a 95% confidence interval that does **not** include the "null value" (null value = 1 for risk ratio or odds ratio and 0 for risk difference). It means that the chance of obtaining such a result by chance is less than 1 in 20 i.e. *unlikely to be due to chance*.

### 2. Could the result be biased?

Bias refers to error. There are many different ways that error can creep into a study. One example is bias with regards to measurement of exposure and outcome. You can imagine that measurement errors could lead to a result that isn't anywhere near the true value.

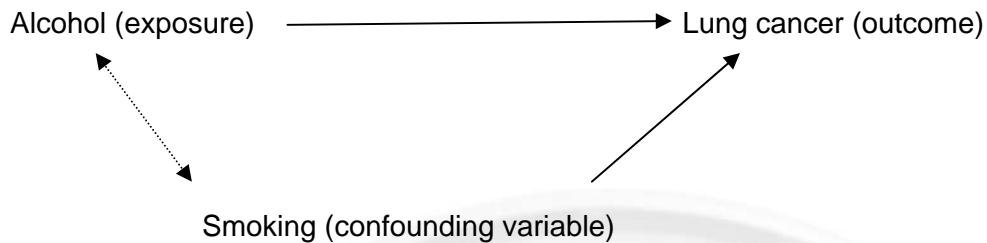
### 3. Could the result be due to confounding?

Confounding (another type of bias) is a slightly tricky concept. Confounding refers to **another variable being responsible for the presumed exposure disease relationship**.

The criteria to satisfy confounding are that

- 1) The confounding variable should be associated with the exposure
- 2) The confounding variable should have a causal role in the outcome independent of the exposure.
- 3) The confounding variable should **not** occur downstream of the outcome.

Let's take my hypothetical study where I proposed that consumption of alcohol is related to the development of lung cancer. Smoking, in this case is a good example of a confounding variable. Smoking is known to cause lung cancer. Smoking is also associated with alcohol consumption. Therefore the presumed effect of alcohol on lung cancer may be due to smoking.



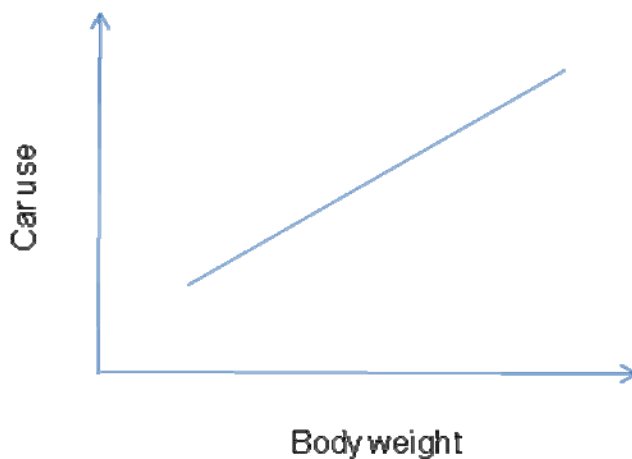
Now that we've considered some of the challenges in study design, let's go on to look at the different types of studies.

## Cross sectional study

**Description:** Measures the prevalence of exposure and disease **in individuals at one period in time**

**Measures:** Scatter graph, correlation coefficient.

**Example:** A survey of patient's body weight and car use showed that body weight was positively associated with car use



**Advantages:** Quick, convenient. Good for formulating hypotheses.

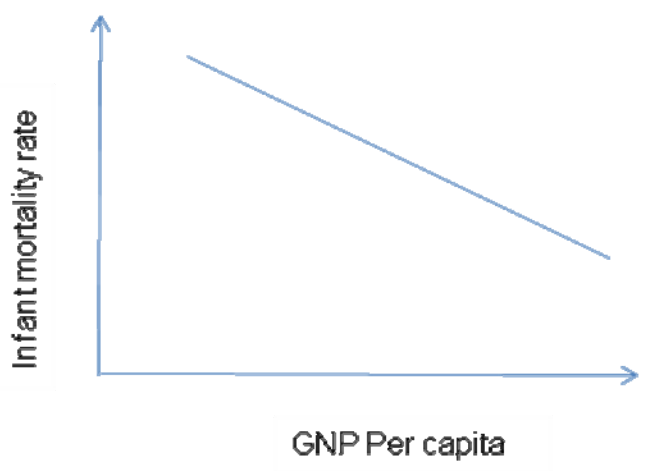
**Disadvantages** Cannot establish temporal relationship between variables e.g. chicken and egg scenario – what causes what?  
May not always account for confounding variables.

# Ecologic study

**Description:** Measures exposure and disease outcomes **at a population level at one period in time**

**Measures:** Scatter graph, correlation coefficient

**Example:** The negative relationship between gross national product per capita and infant mortality rate across countries.



**Advantages:** Quick and convenient to do. Good for formulating hypotheses.

**Disadvantages:** Similar to cross sectional study. In addition, be aware of the **ecologic fallacy**- relationships between variables that are true on population level **cannot** be assumed to be true on an individual level.

## (Prospective) Cohort study

**Description:** Subjects should be without the disease at the beginning. The study follows those who have been exposed to a particular variable and those that haven't (NB you don't have control over the exposure unlike a randomised controlled trial)  
At the end of the study period, measure the incidence of disease in the two groups and compare them

**Example:** Smoking and incidence of myocardial infarction (MI)

	MI	No MI	Total
Smoker	20	180	200
Non smoker	5	95	100

**Measure:** Risk Ratio or Risk difference  
 Risk Ratio = Incidence in exposed/Incidence in unexposed  
 = (20/200) / (5/100)  
 = 2  
*Those who smoke, have twice the risk of developing an MI compared to those who don't smoke*  
 Risk difference= Incidence in exposed – incidence in unexposed  
 = (20/200) - (5/100)  
 = 5 per 100  
*There is an excess risk of MI of 5% in the smokers compared to those who don't smoke.*

**Advantages:** Good study for examining temporal relationship between exposure and outcome.  
 Can adjust for most confounding variables if they are accounted for in the study design.

**Disadvantages:** Require long follow up period, and not so good for rare outcomes.

## Case-control study

**Description:** Diseased and non diseased individual are identified. Exposure is retrospectively ascertained for all individuals. Look at the odds of exposure in each of the diseased and non diseased groups.

**Example:** Investigating whether a childhood disease x is due to an exposure to a drug *in utero*

	Childhood disease x (cases)	Non diseased (controls)
Drug in <i>utero</i>	100	10
No exposure	50	50

**Measure:** **Odds ratio**  
 Odds ratio = Odds of exposure in diseased group/ odds of exposure in non diseased  
 = (100/50) / (10/50)  
 =10

*The odds of someone developing disease x after exposure to drug in utero is 10 times more than if they had not been exposed.*

**Advantages:** A good study design to use for rare outcomes (e.g. cancer)

**Disadvantages:** You are unable to tell the absolute risk of a disease in an exposed or unexposed group. You can only make a comparative statement such as the odds ratio.  
 Prone to recall bias (e.g. diseased subjects may recall exposure more favourably than non diseased subjects).

# Randomized Controlled trial

- Description:** Participants randomized to two groups- exposed and non exposed (note that exposure is assigned in this study unlike the cohort study). Measure disease in the two groups during the study or at the end of the study and compare.  
**Double blind-** refers to the fact that both the researcher and participant are blind to the treatment that they are getting.  
**Intention to treat analysis** – refers to the fact that data of individuals are analyzed on the assumption that they stayed in their respective exposure groups (e.g. drug or placebo).
- Example:** Trial comparing drug A versus placebo in prevention of colorectal cancer.
- Measures:** Risk difference or risk ratio  
(see calculation for prospective cohort study)
- Advantages:** The Gold standard of study designs. Randomization minimizes confounding.
- Disadvantages:** Ethical limitations- very difficult to randomize in circumstances where you know there will be harm done to the control group. Also difficult to do randomised controlled trials for invasive procedures.

## About the Author

**Reuben Arasaratnam** is a currently a Foundation Year 2 Doctor working in the United Kingdom, with a Masters in Public Health. He is helping to write a series of guides for MedicalEducator. Medical Educator would like to thank him for his contribution.