Strengthening the Occupational Health Expertise and Scientific Performance of Public Health Institution of Turkey

Study protocol Lesson 3 : Methods

A 2.3 ppt; third presentation on day 2 of training A

Part of the Chapter Methods of the protocol

Learning Objectives

1. You are able to select an appropriate study design

2. You are able to define your target population and study population

3. You are able to estimate the statistical power of your study

Summary of presentation

- Study design
- Study population
- Statistical power

Study protocol Part 3

The strategy of the investigation, the target population

What is epidemiology?

Epidemiology is the science and practice that analyses the occurrence of **diseases** in different **groups** of persons and the **why**.

Epidemiologists are the detectives of the diseases

Epidemiology
-Formal definition-

Epidemiologists study

- Factors that have an influence on **health:**

  – The start of a **disease**

  – The diagnosis

  – The prognosis (course)

  – The consequences

- **Prevention and treatment** e.g. effectiveness of interventions

Two types of epidemiological studies

Observational study

e.g. risk on ebola for health care workers in a region in a certain period of time

Intervention study

e.g. evaluation of effectiveness of specific guidelines (plus PPE, instruction and supervision).

Designs of a descriptive study

Cross-sectional study

Case-control study

Cohort study

Case study

Qualitative study

Can you give a definition and an example?

Cross-sectional study

- Not taking into account *time course*
- Exposure and prevalence of disease measured at same time
- Useful
  - first information about **association** between exposure and disease
  - for development of surveillance programs
- Not permitted to conclude that the exposure preceded the disease or vice versa. What came first : the egg or the chicken?
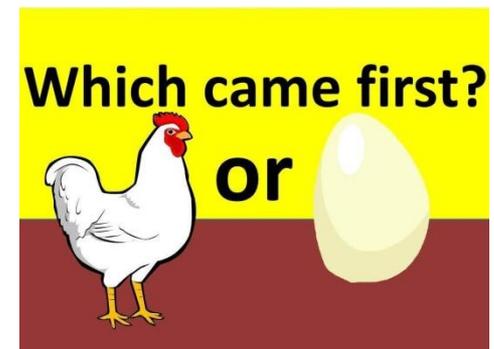
Cross-sectional study: (dis)advantages

- Easy to complete
- Quick, cheap („a bargain")
- First impression of the problem
- You can study various diseases/health complaints and exposure at once
- Adequate for chronic diseases that are not fatal such as allergies, asthma, COPD, chronic low back pain, chronic stress complaints
- Useful for more commons diseases

Cross-sectional study: (dis)advantages

- Not useful for rare diseases

- Not useful for diseases with a short duration

- Not appropriate for causal investigations

- But you can find an association and check this in a new or existing other studies

- Take care of information bias and selection bias

Your protocol

- Explain briefly why chosen for a cross-sectional study or for another design.

- You must also define where the study will be carried out.

- Other details: see the Manual.

- This subchapter is mostly short (maximum a half page).

Study plan or Methods

Type and design of the study?

Population?

Instruments?

Bias, quality (errors)?

Ethics?

Statistical plan?

Strong and weak points?

# Studies should reflect the real situation

Target population and Study population


Target population

Population that is the study object

All workers/patients that the researcher is interested in

Target population and Study population
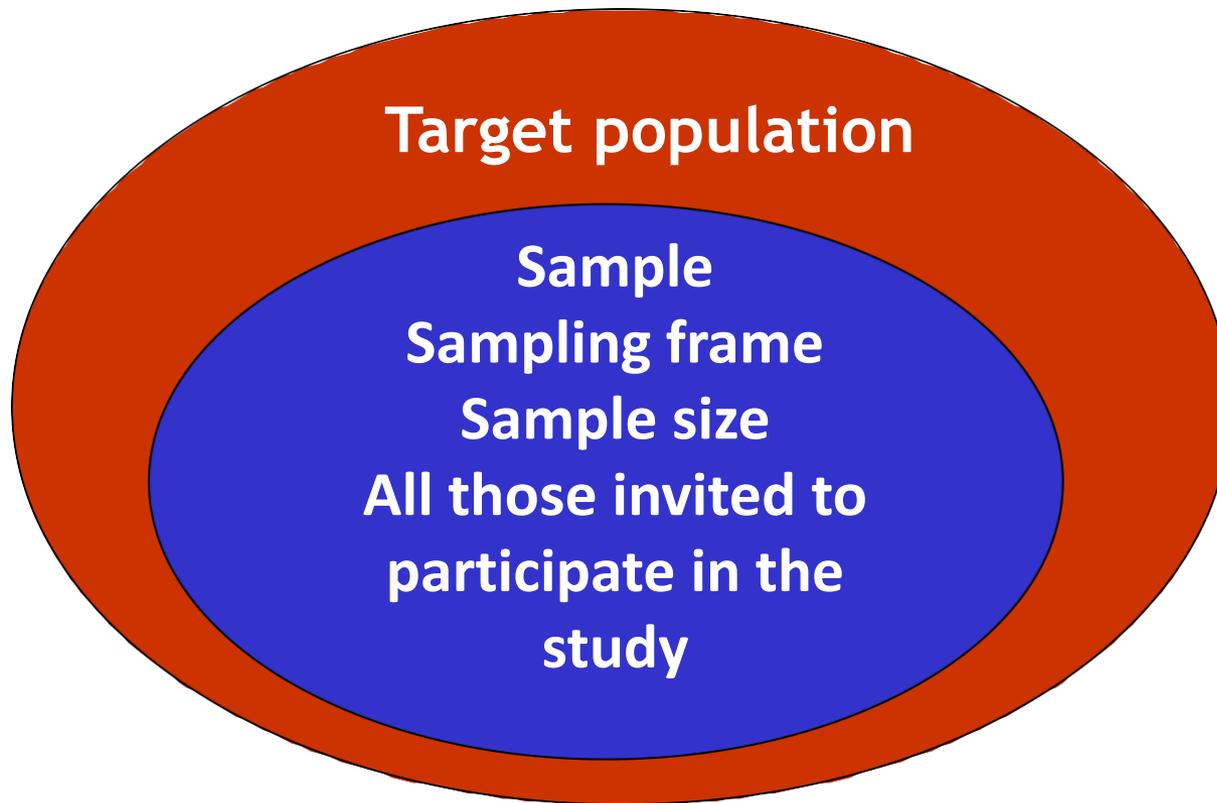
Target population

**Sample
Sampling frame
Sample size
All those invited to participate in the study**

Target population and Study population

Target population

Sample
Sampling frame
Sample size n=900
Number of responders
plus non-responders is
e.g. 600+300=900

Target population and Study population

Target population

Sample

Study population is the sample studied

All who participate
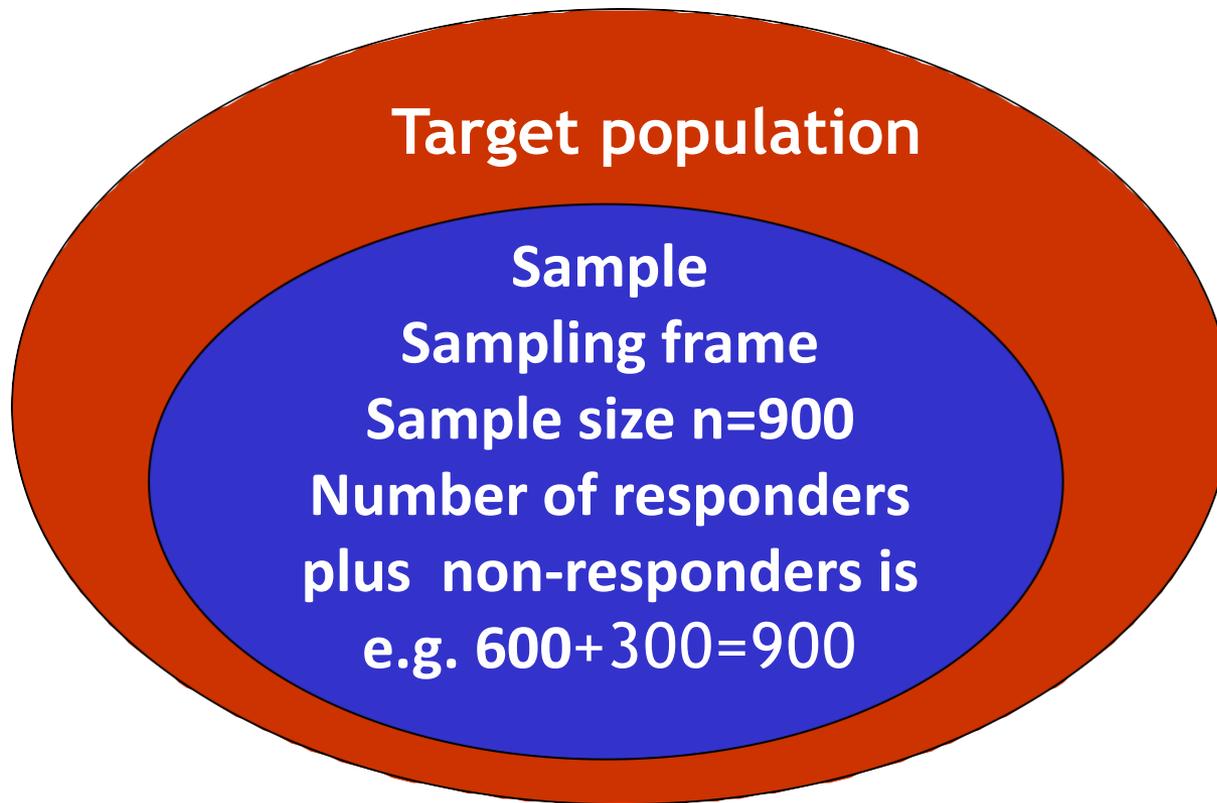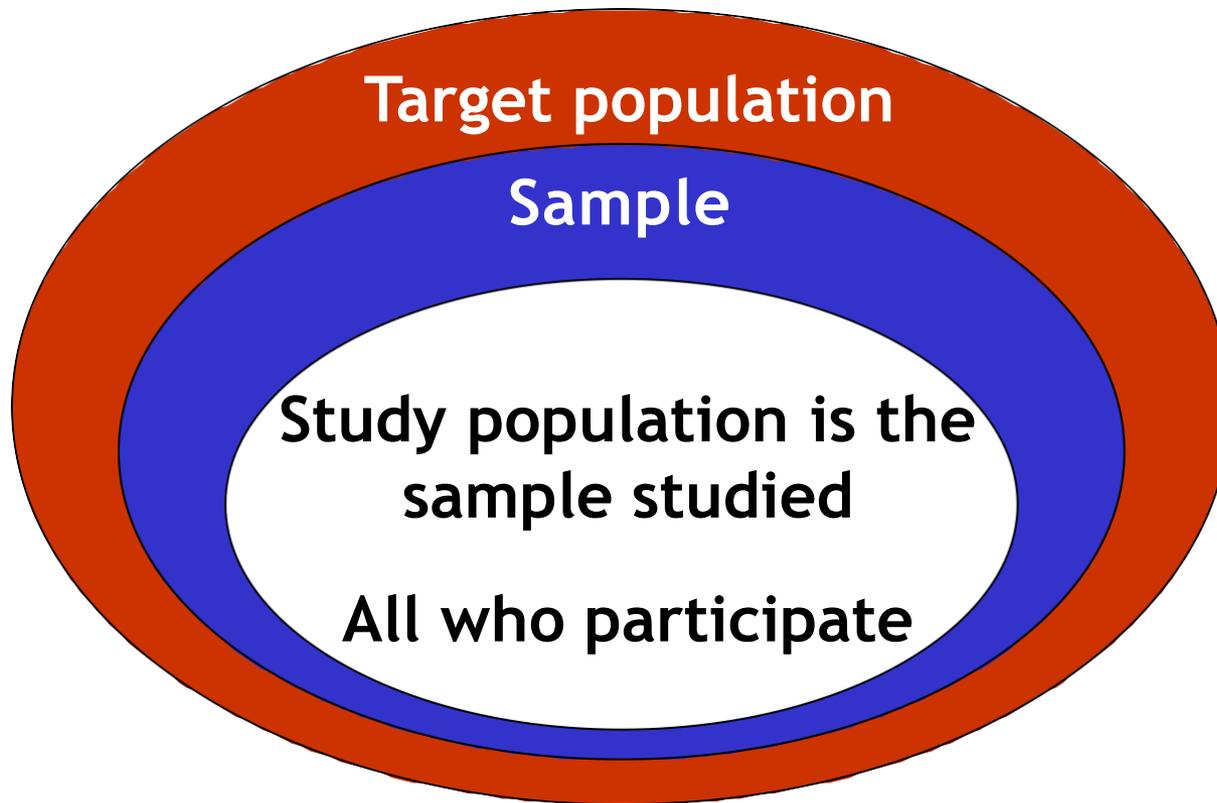
Example:
# Health effects in welders in Turkish metal industry

Target population

All welders in Turkish
metal industry:
formal and informal
workers

Target population and Study population

Target population

Sample

(Random) sample
of welders
N=900

Target population and Study population

Target population

Sample

Study population

600 welders

(response rate

66%)
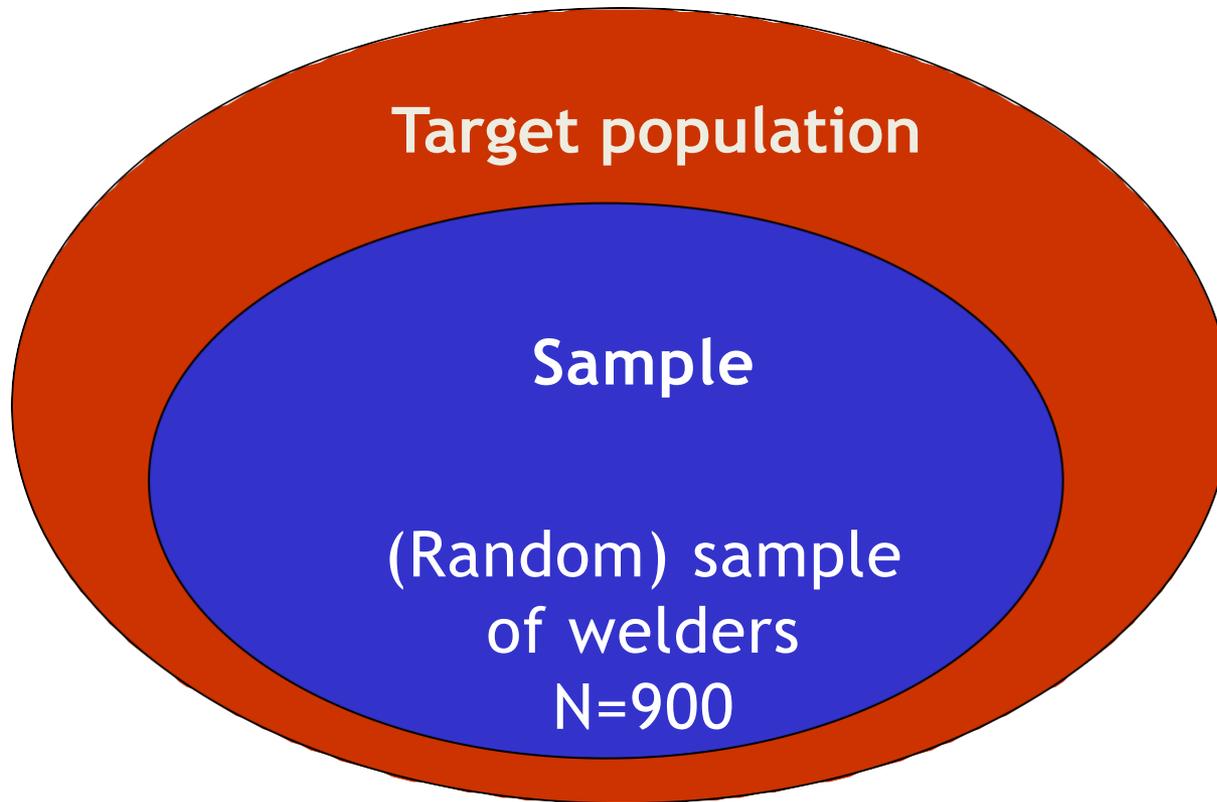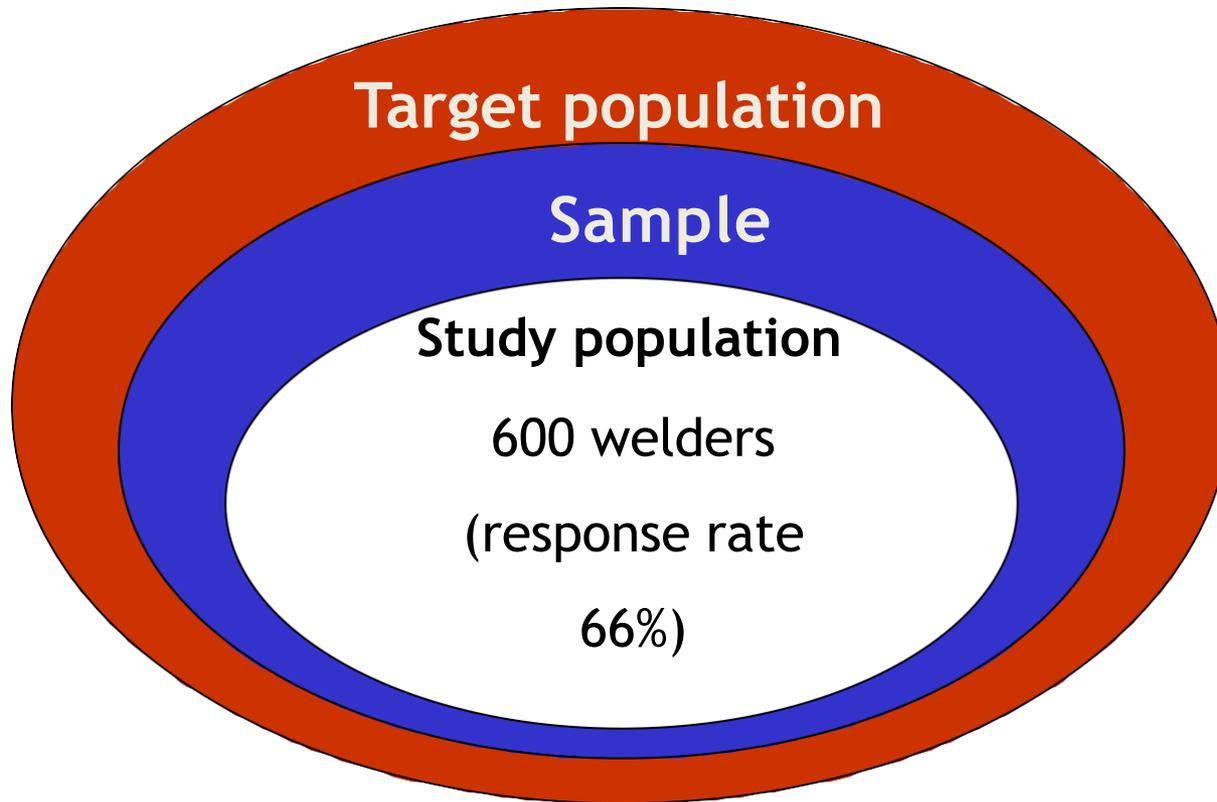
Validity of a study

Internal validity

      The association found reflects in a reliable way the association between exposure and disease in the study population

      Sources for errrors (bias):

            Selection bias, information bias, confounder bias

External validity

      The association found is correct for the target population. A related term is representativeness. It is important to select an appropriate sample!

Your target population
- Examples -

- All workers in the informal sector, working for hotels and restaurants in Antalya.

- All outsourced (subcontractor) workers in a large company.

- All patients at the respiratory diseases department in the main hospital in one city in 2017.

- All patients diagnosed with mesothelioma in a Occupational Diseases hospital or included in a national mesothemioma register in 2007-2016.

Your sample: option 1

When there is a complete list of your target population:

- You can take a random sample from this list, e.g. with random numbers

- You can invite all members of this sample to participate in your study

Invitation letter, informed consent, visit, etc.

Important

Application for the Ethical or institutional committee for revision of ethical aspects!

Confidentiality, etc.

Improve the response rate

- Why?

  -> Reduce selection bias

Examples?
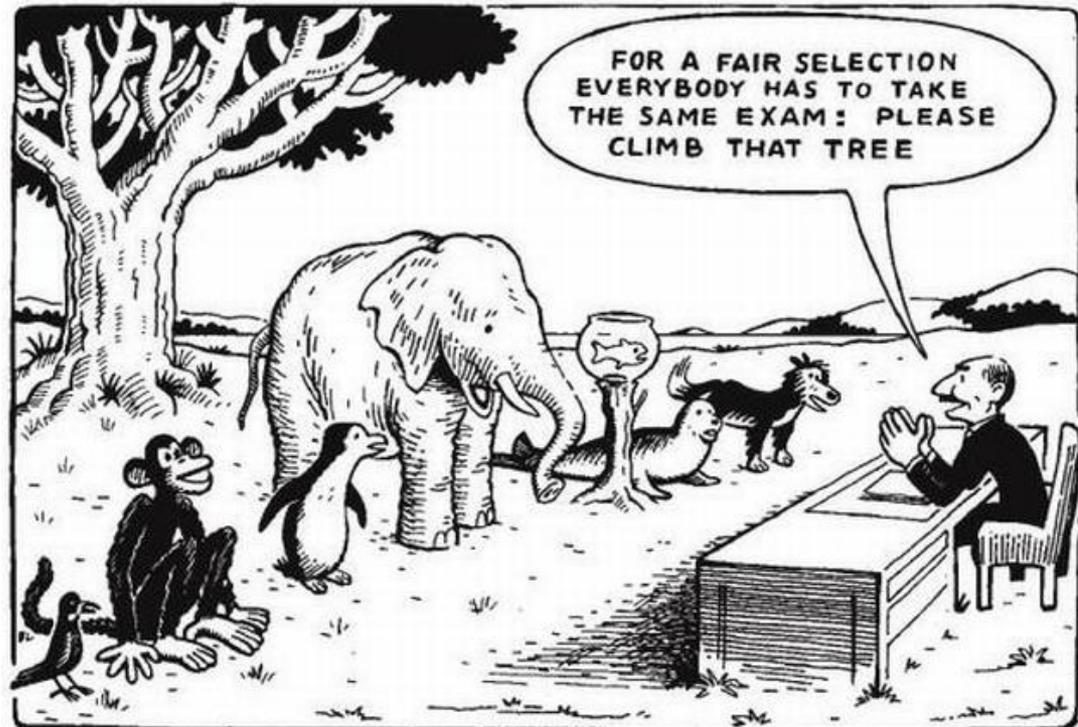
Selection bias

How is selection bias produced?

Example on respiratory symptoms in cement industry

- You want to compare exposed (mostly informal) workers and non-exposed (mostly formal) workers in this industry.
- Finding: respiratory symptoms in non-exposed workers are slightly higher than in exposed workers.
- Smoking habits, ages and gender are equal.
- Is this result reflecting reality? Possible selection in participation?

Selection bias

Target pop. symptoms
Exposed vs Control
60% vs 46%

Exposed
(informal)
workers
n=80
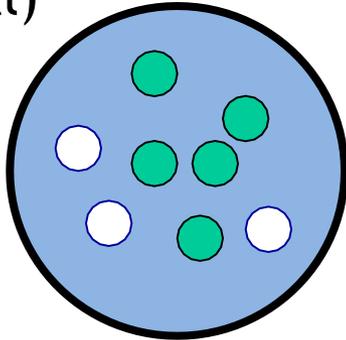
Exposed
(informal)
workers
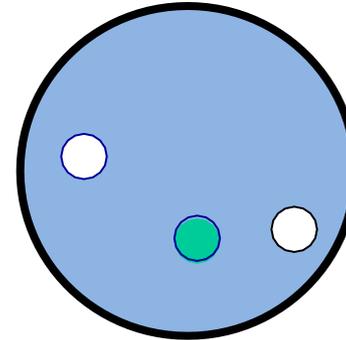
Study pop. symptoms
Exposed vs Control
33% vs. 40%

Not-
exposed
(formal)
workers
n=80
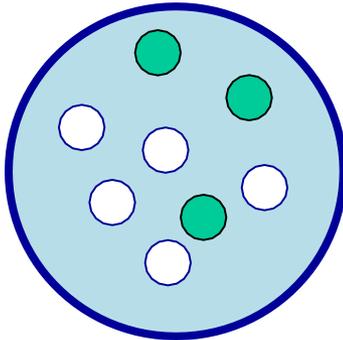
Not-
exposed
(formal)
workers
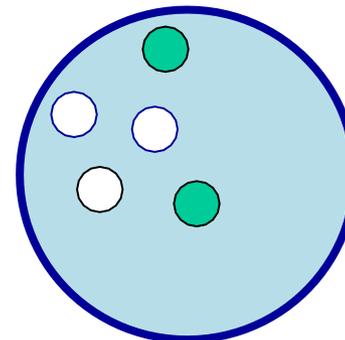
○ Healthy
● Not healthy

Response rates = …. ?
What might have
caused selection bias?

Target population ⟶ Study population

Reduce selection bias

- Neutral invitation letter

  "Study about the working conditions" without mentioning specific diseases or the informal sector

- Maximize the response rate

  - Reminders

  - Short questionnaire

- Better: give a reward: a present, money, a voucher

Representativeness of your study population: no 'us'

It is not good to select your friends or neighbors as study population (convenience sample) or to visit workers only one time e.g. Wednesday 11 a.m.

Statistical power of the study
How many workers/patients have to be included?

- When you analyse a too small number of workers in order to demonstrate a significant difference, all the work invested for the study may be in vain because the results do not help anything

- When you analyse a too high number, you are using resources, your time, the time of the participants; al this is not needed to demonstrate the difference.

- Both cases would be not ethical

The whole population may be large, subgroups may be small for analysis

Table. Subgroups in Spanish working population, from the VI National Working Conditions questionnaire 2007.
Total number of participating workers: 11.054 workers (study population)

Emergent collectives of workers	N workers
Self-employed with employees	527
Self-employed without personnel	1.421
Employees with a salary (employed) without social security protection	456
Migrant workers (foreign workers)	1.469
Workers on sick leave for different reasons	388
Domestic workers (cleaning, child care)	246

Statistical power

There is nothing cast in stone regarding the appropriate level of **statistical power**, but Cohen (1988) reasoned that studies should be designed in such a way that they have an 80% probability of detecting an effect when there is an effect there to be detected.

To put it another way, studies should have no more than a 20% probability of making a Type II error (recall that power = 1 – β).

https://effectsizefaq.com/2010/05/31/what-is-an-ideal-level-of-statistical-power/

Statistical power of the study
How many workers/patients have to be included?

When you have two samples
1. Medium-size companies (total N=**100** workers)
2. Small companies (total N=**100** workers),

An **alfa level of 5 %** (type 1 error) and

a. A prevalence of a symptom or working condition of **30**% in (1) and **50%** in (2), the statistical power is **83%**;

b. A prevalence of a symptom or working condition of **10%** in (1) and **30%** in (2), the statistical power is **95%**;

c. A prevalence of a symptom or working condition of **10%** in (1) and **20%** in (2), the statistical power is **51%**.

Statistical power of the study
How many workers/patients have to be included?

When you have two samples
1. Medium-size companies (total N=**200** workers)
2. Small companies (total N=**200** workers),

An **alfa level of 5 %** (type 1 error) and

a. A prevalence of a symptom or working condition of **30%** in (1) and **50%** in (2), the statistical power is **98%**;

b. A prevalence of a symptom or working condition of **10%** in (1) and **30%** in (2), the statistical power is **100%**;

c. A prevalence of a symptom or working condition of **10%** in (1) and **20%** in (2), the statistical power is **80%**.

Statistical power of the study
How many workers/patients have to be included?

When you have two samples
1. Medium-size companies (total N=**250** workers)
2. Small companies (total N=**250** workers),

An **alfa level of 5 %** (type 1 error) and

a. A prevalence of a symptom or working condition of **30**% in (1) and **50%** in (2), the statistical power is **100%**;

b. A prevalence of a symptom or working condition of **10%** in (1) and **30%** in (2), the statistical power is **100%**;

c. A prevalence of a symptom or working condition of **10%** in (1) and **20%** in (2), the statistical power is **88%**.

Statistical power of the study
How many workers/patients have to be included?

This depends largely of the prevalence of the disease/health complaints and the difference between both populations that you expect and consider relevant.

You have to consider if a difference in working conditions of e.g. only 10% could be relevant to induce a change in disease frequency after the intervention.

Sample size calculator

http://www.openepi.com/SampleSize/SSMean.htm

Statistical power of the study Commitment for your project

- When you do a questionnaire study, consider to include at least 100-200 workers in each of your subgroups.
- You need at least two subgroups for a comparison, mostly Exposed vs. Not exposed workers, sometimes Informal vs. Formal workers in the same company/branch.
- In general, do not select more than 2 subgroups.
- If you perform your study in many small companies or in self-employed, you have to identify 2 groups for the analysis.

Attention!

The group without exposure has to be really (as much as possible) **without exposure**.

Example

When you want to evaluate the prevalence of hearing impairment in carpenters, it makes no sense to compare the audiometric results with those of the population of 'all workers in the construction industry'.

Why?

Attention!

- When your response is 50%, you should invite twice the number needed for the power.

- Do not forget to mention how many persons you plan to invite to participate in the study!

- Only in this way you will be able to evaluate the response rate.

Your protocol

- Describe briefly (½ -1 page) your
  - Target population,
  - Your sample,
  - Your study population.
- Explain why you have decided to study this target population.
- Explain why you decided to select this sample, how to approach the population, what to do to increase the response.
- Do not forget to indicate the number of participants that you will examine (look at the power).

For more details: see the Manual.

Workshop 3

Work in small groups:

- Develop a study strategy

- Define the target and study population

- Look at the power of the study