# Study design

Prof. Giuseppe Verlato

Unit of Epidemiology & Medical Statistics
Dept. of Diagnostics & Public Health
University of Verona

## Epidemiology

| Observational | | | | Experimental | | |
|---|---|---|---|---|---|---|
| **Descriptive** | **Analytic** | | | **Evaluative** | | **Clinical epidemiology** |
| Cross-sectional *(trasversale)* | Cohort study | Case-control study | Echologic study | Field trial (Intervention on individuals) | Community intervention trial | Controlled (randomized) clinical trial |

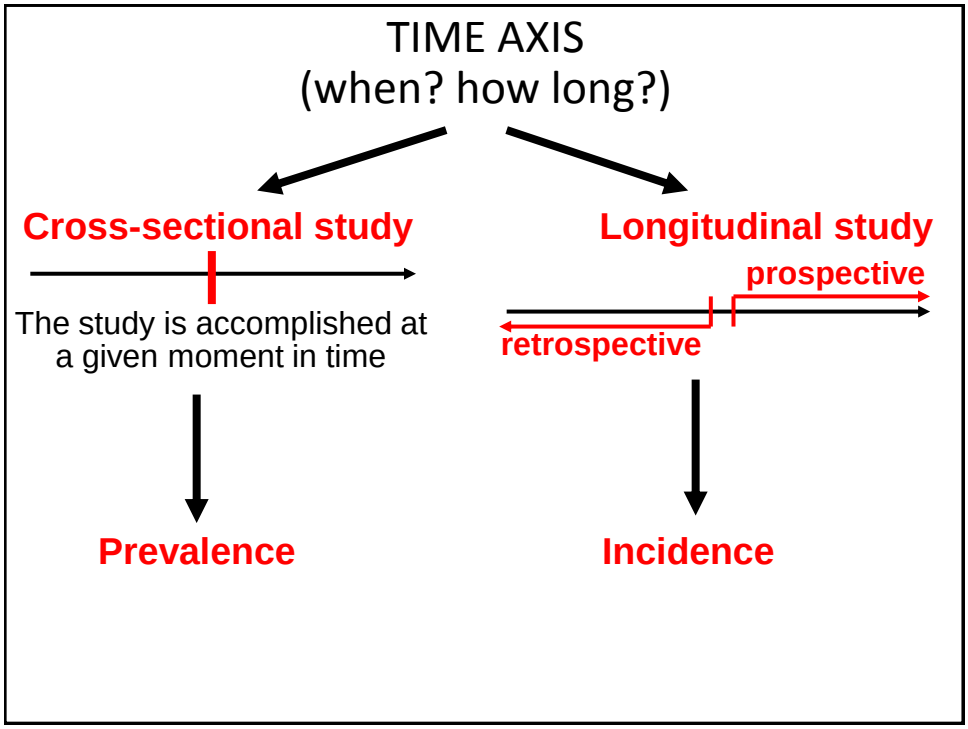
| EXPERIMENT | PLANNED OBSERVATION |
|---|---|
| Researchers actively modify the course of events | Researchers just observe the course of events, without attempting to modify it |
| Only **positive perturbations** can be applied:<br>1) Preventive interventions, such as adding fluorine to tap water, or iodine to salt<br>2) Therapeutic measures (early thrombolysis in myocardial infarction, segmental vs total mastectomy)<br>3) Rehabilitation interventions | Also **etiologic factors** with **deleterious** health effects can be studied:<br>1) wrong lifestyle (smoking, excessive alcohol intake)<br>2) environmental situation (Chernobyl) |
| **RANDOMIZATION** | **SELF-SELECTION** |
| Participants are randomly assigned to different treatments<br>↓<br>Other risk factors (potential confounders) are balanced among groups | Potential confounders are not eliminated. For instance, it could be hypothesized that:<br>Unknown genes → Craving for smoking<br>Unknown genes → Increased risk of lung cancer |

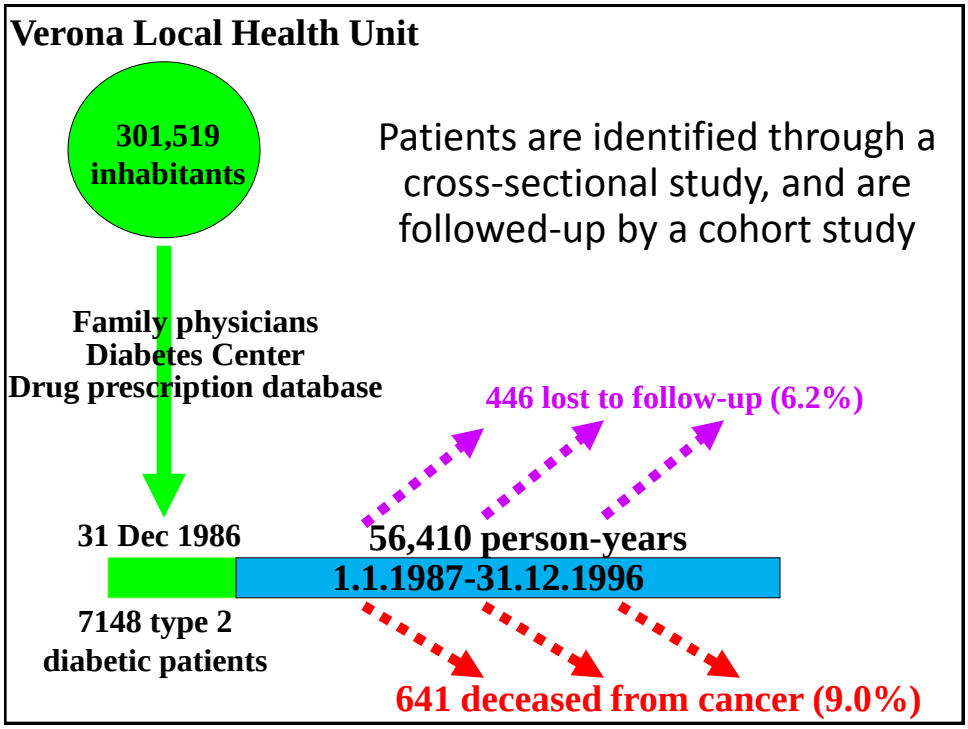
## Observational studies

- If a diagnostic procedure (for example TC-scan), which does not modify the course of disease, is added to normal clinical practice

According to epidemiologists → This is an observational study

According to the Italian law (G.U. **31/03/2008** – AIFA - Determinazione 20 marzo 2008) → The study is NO LONGER observational


TIME AXIS
(when? how long?)
Cross-sectional study
The study is accomplished at a given moment in time
Prevalence
Longitudinal study
prospective
retrospective
Incidence

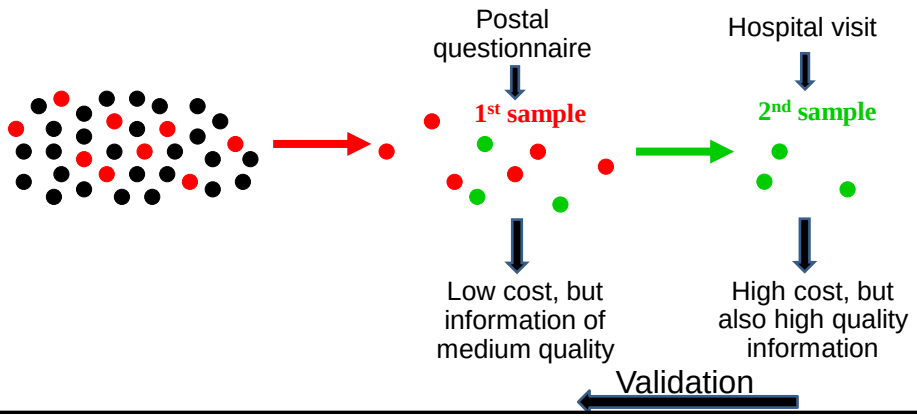

Verona Local Health Unit
301,519 inhabitants
Patients are identified through a cross-sectional study, and are followed-up by a cohort study
Family physicians
Diabetes Center
Drug prescription database
446 lost to follow-up (6.2%)
31 Dec 1986
56,410 person-years
1.1.1987-31.12.1996
7148 type 2 diabetic patients
641 deceased from cancer (9.0%)

# CROSS-SECTIONAL STUDY

- It allows to estimate prevalence of diseases
- It is suited to study chronic-degenerative diseases (diabetes, hypertension, osteoarthrosis):
  long duration → high prevalence
  (Prevalence ≈ Incidence*duration)
- It does not allow to infer cause-effect relation, but only to describe association between diseases and determinants
- It can be based on samples or on the entire population (exhaustive studies, like censuses)
- If the cross-sectional study is based on samples, sampling should be correct (probability sampling)

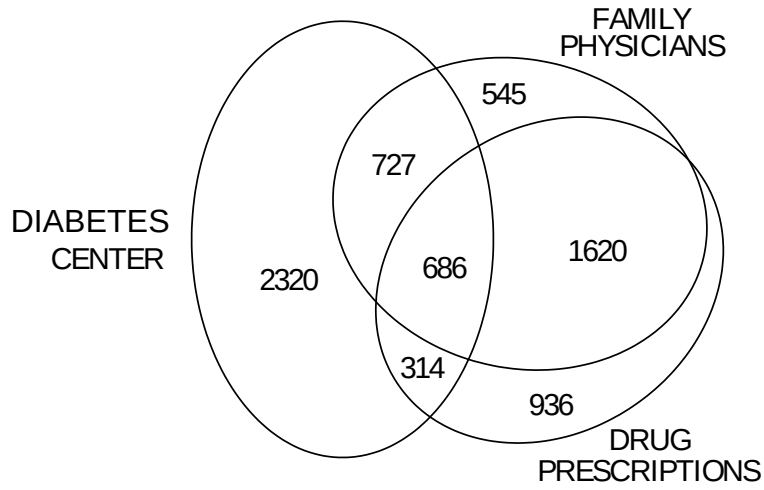
## Example of a SAMPLE study: European Community Respiratory Health Survey

**TWO PHASE SAMPLING: A 1st sample is randomly drawn to undergo cheaper investigations (postal questionnaire). From the 1st sample a 2nd sample is randomly drawn to undergo more expensive investigations (clinical visits, lab tests).**



## Example of EXHAUSTIVE study: the Verona Diabetes Study

7,148 type 2 diabetic patients identified through three different sources




**Verlato and Muggeo, Diabetes Care, 2000**

## Cross-sectional studies do not provide information on causal links

**In a cross-sectional study [Verlato, Int Arch Allergy Immunol 2016]:**

**P(asthma/male never smokers) = 5.5%**

**P(asthma/male current smokers) = 4.2%**

**What do these findings suggest? Does current smoking protect men against asthma ?!?!?!?**

**No, it doesn't. "Asthma in childhood reduces smoking initiation during the subsequent teenage in men" [Verlato, J Adolescent Health 2011].**

# Cohort study

**A fixed population (cohort)** is a group of individuals:

1) identified as they experienced a common event at time zero ($t_0$, beginning of the study)
2) Followed-up over time

**Inclusion criteria:**

1) Subjects must NOT have the disease under study
2) Information on subjects' exposure to risk factors must be available

# Cohort study

**Exit criteria:**

1) Occurrence of the event under study (death, onset of disease, onset of complications, ...)
2) End of follow-up (for instance 31° of December 1996 in the Verona Diabetes Study)
3) Loss to follow-up (it should be kept under 10%)
4) Participant's death if death is not the event under study (problem of competitive risk can arise)

## Example of a cohort study

1) A study aims at investigating the relation between smoking habits and myocardial ischemia
2) At the beginning of the study, a cohort of subjects without myocardial ischemia is enrolled
3) Participants' smoking habits are assessed
4) Subjects are followed-up over time in order to record new cases of myocardial ischemia among smokers and non-smokers

## Cohort studies - Pros

1) … allow to **reduce the sources of bias**
2) … allow to measure the **incidence** of diseases
3) Hence they allow to estimate **all measures of association** (RR, RD, etiologic fraction)
4) … allow to collect information on **early phases** of diseases
5) … represent the most efficient design when **the exposure is rare**; cohort studies are ongoing in Černobyl' and Seveso.

More than 4,000 cases of thyroid tumors were observed among people aged <18 years in Chernobyl in the 25 years following the disaster.

## Cohort studies - cons

1) Cohort studies, being usually prospective, are **complex, long-lasting, expensive**
2) They are **not suited** for **rare** diseases or diseases **with long latency** (pleural mesothelioma)
3) During the follow-up diagnostic tools and/or treatments can improve, posing **ethical dilemmas**
4) Cohort studies become less reliable when **more than 10% of subjects are lost to follow-up.** It can be difficult to **trace** these subjects, who have usually moved to other areas.

## Case-control studies

1) Enrolled subjects are classified according to disease status as cases (diseased) and controls (healthy) at the beginning of the study.
2) Subjects' exposure to risk factors in the past are ascertained.

From a practical perspective:

1) All cases existing in a given place at a given period are identified
2) One to ten controls per case are drawn from the same population as the cases. Controls should be as similar as possible to cases as regards sex, age, site of residence, job, …, but should be disease-free.
3) Past exposures of cases and controls are ascertained.

# Example of case-control study

Eight women aged 15-22 years living in Boston were diagnosed with adenocarcinoma of the vagina from 1966 to 1969.

UNUSUAL CLUSTER: "Cancer of the vagina is rare, occurring usually as epidermoid carcinoma in women over the age of 50 years."

DISEASE LOOKING FOR A CAUSE

RARE DISEASE → CASE-CONTROL STUDY

NEED TO URGENTLY FIND THE CAUSE

8 CASES ⟷ 32 CONTROLS

1 CASE → 4 controls with the same age, born in the same hospital ...

## Past history

- Maternal age = no difference
- Maternal smoking = no difference
- Bleeding in pregnancy in 38% of cases and in 3% of controls
- Prior pregnancy loss in 75% of cases and in 16% of controls
- treated with diethylstilbestrol started during the first trimester of pregnancy in 88% of cases and in none of the controls

THE CASE-CONTROL STUDY IDENTIFIED THE CAUSE OF THE DISEASE

DIETHYLSTILBESTROL IN PREGNANCY ⇒ ADENOCARCINOMA OF THE VAGINA 15-20 YEARS AFTER UTERINE EXPOSURE

Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the Vagina — Association of Maternal Stilbestrol Therapy with Tumor Appearance in Young Women. N Engl J Med 1971; 284:878-881.

## Case-control studies - advantages

1) Case-control studies are useful to investigate **diseases looking for a cause.**
2) … are suited for **rare diseases with long latency.**
3) … are **quick and cheap.**
4) … do not require **large samples.**
5) … allow to use **pre-existing** information.

## Case-control studies - advantages

«The case control study design can be considered a more efficient form of the follow up design, in which the cases are the same as those that would be included in a follow-up study and the controls provide a fast and inexpensive means of inferring the distribution of the exposure in the population that gave rise to the cases".

Rothman 1986. Modern epidemiology

In Italian: «Se condotti in modo appropriato, gli studi caso-controllo consentono di confrontare tutti i casi, insorti in un determinato periodo, con un campione dei soggetti sani tratti dalla stessa popolazione.»

## Case-control studies - disadvantages

1) … are **very prone to bias**
2) exposure data are **secondhand**
3) … do not provide any measure of disease frequency (prevalence, incidence), hence they allow to compute only the **odds ratio.**
4) … are not recommended when **case identificatio is incomplete.**
5) … are **not suited** when **the exposure is rare.**
6) for all these reasons they are often defined as **«quick and dirty»** (veloci e sporchi).

## Examples of bias in case-control studies

Physical exercise → myocardial infarction !!!

Sportsmen → survive myocardial infarction → are recruited as cases

Sedentary people → do not survive myocardial infarction → are not recruited as cases

**Selection bias = prevalent cases are recruited rather than incident cases**

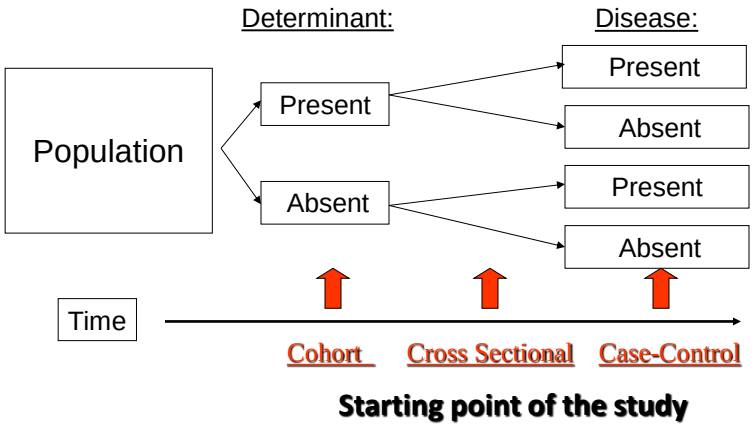
Everything → fetal malformation !!!

Mothers of children with congenital malformation → remember all events which have occurred during pregnancy

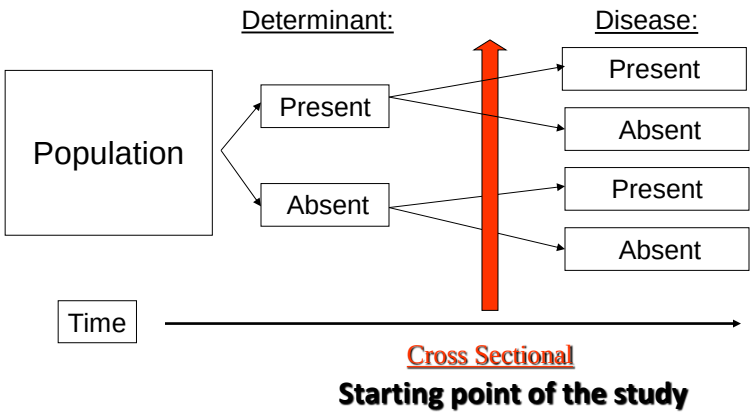
Mothers of healthy children → forget many events occurring during pregnancy

**Recall bias**

## Comparison of different study designs

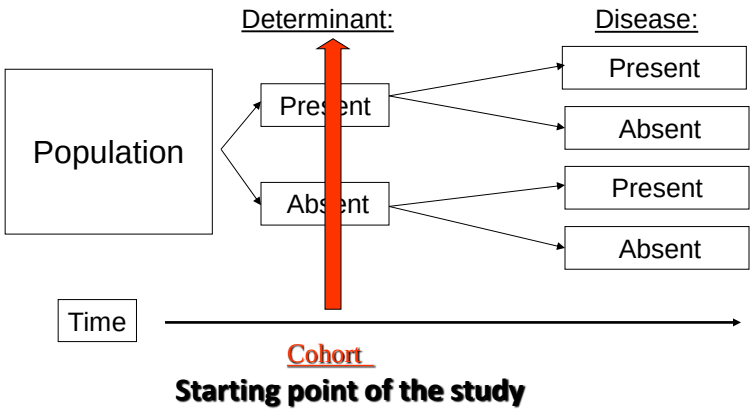


# Cross-sectional studies



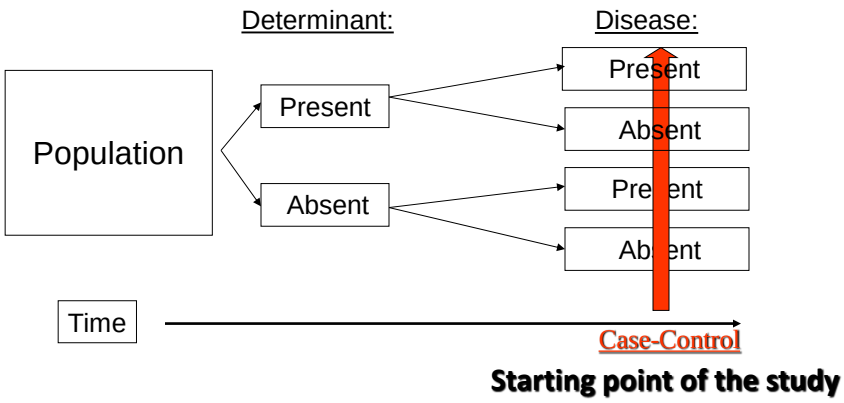

Information on time relation between potential determinants (often risk factors) and disease is lacking

# Cohort studies




The study goes from the cause to the effect

# Case-control studies




The study goes from the effect to the cause

## ECOLOGICAL STUDIES

**Ecological studies investigate spatial and/or temporal relation between a determinant and a measure of occurrence,**

**at population level rather than at individual level (oikos = home).**

**Ecological studies can deal with Local Health Unit, administrative regions, municipalities.**

## EXAMPLES of ECOLOGICAL STUDIES

**In Naples and Barcelona, epidemics of asthma were recorded when ships full of soy were unloaded.**

**A TEMPORAL relation exists between air pollutant concentrations and mortality or hospitalization from respiratory diseases.**

**GEOGRAPHIC RELATION: In England suicide rate is lower in towns having a branch of the Samaritans, an organization offering help to suicidal or equally desperate people.**