# EPIDEMIOLOGICAL STUDIES INCLUDING BIAS

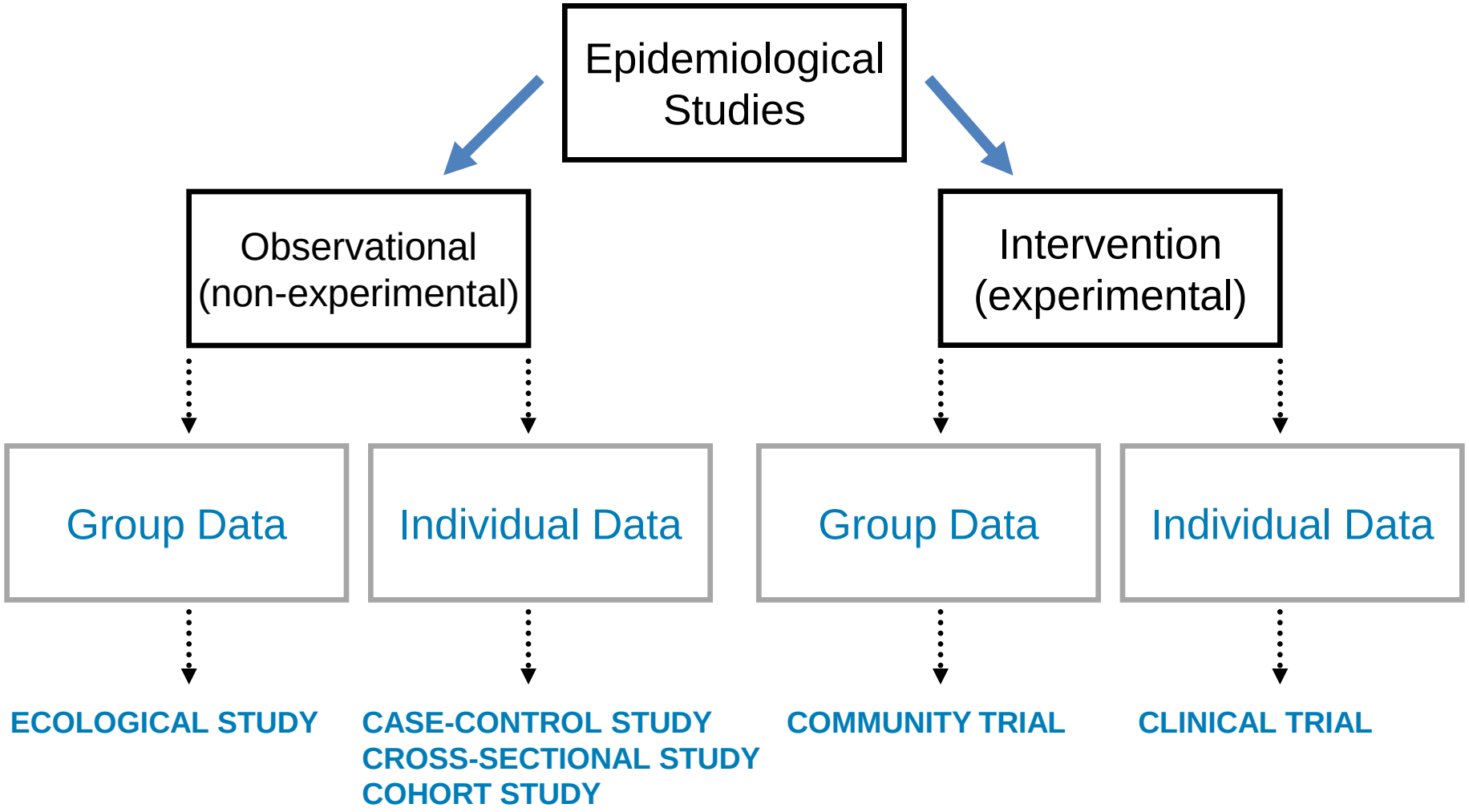
Olivier Bruyère, PhD

Professor of clinical epidemiology and Professor of geriatric rehabilitation, Dep. of Public Health, Epidemiology and Health Economics, CHU - Sart Tilman, Liège, Belgium

# EPIDEMIOLOGICAL STUDIES

- Introduction
- Bias and confounding
- Cohort studies
- Case-control studies

# STUDY DESIGN OVERVIEW

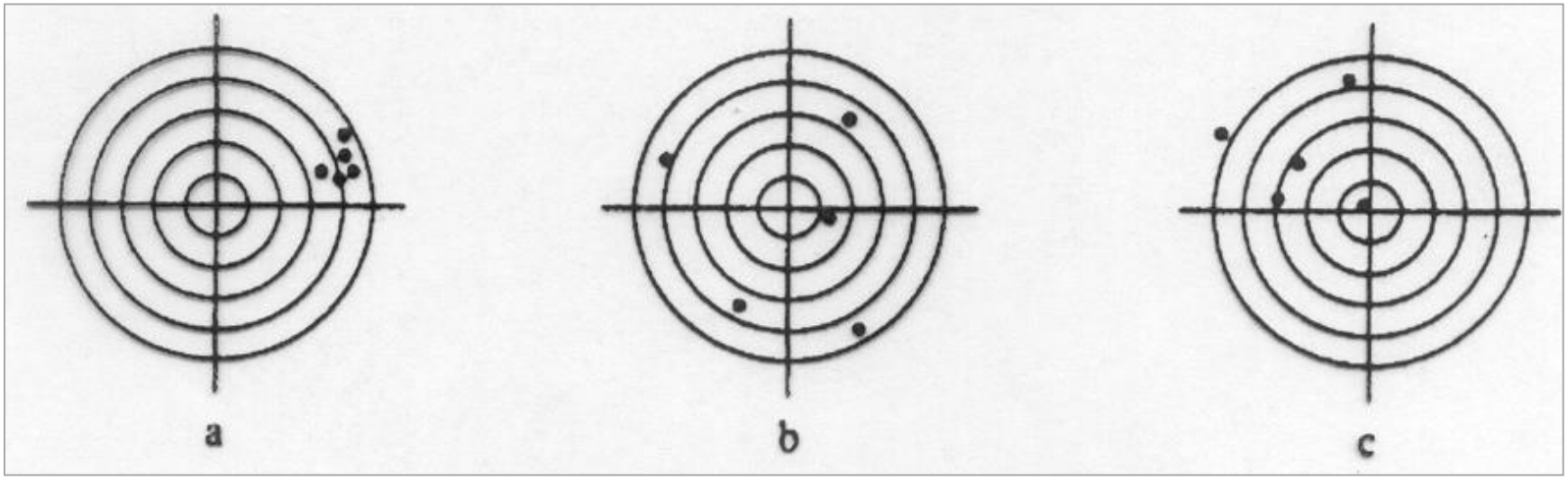


# 1. BIAS AND CONFOUNDING

1. Definition
2. Selection bias
3. Information bias
4. Confounding factor

# 1.1. DEFINITION

Random and systematic errors



# 1.1. DEFINITION

## Bias and chance

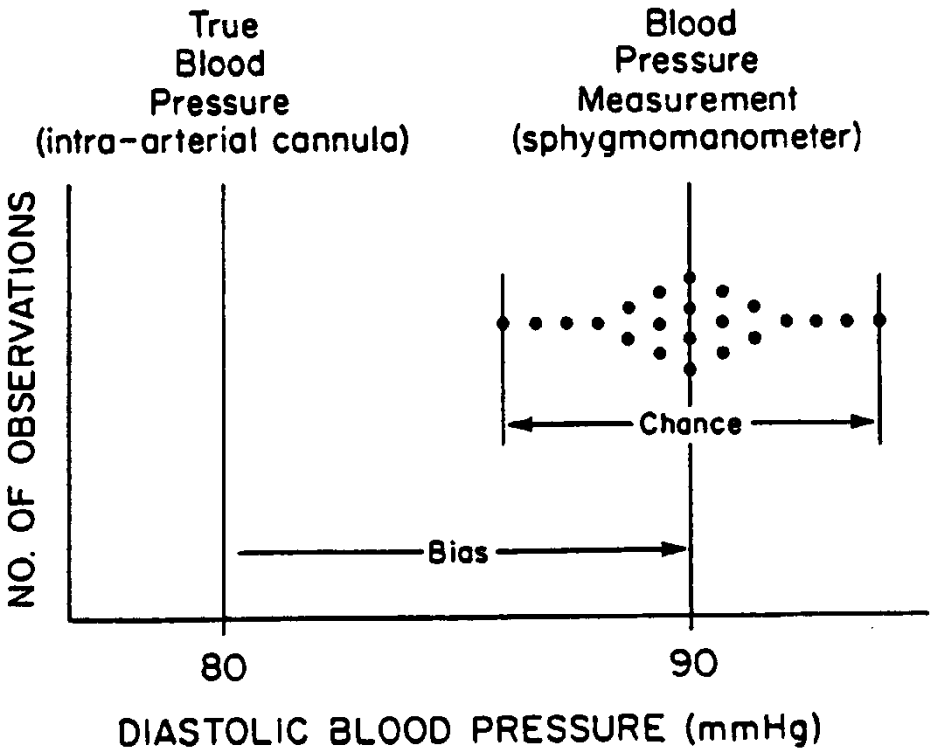
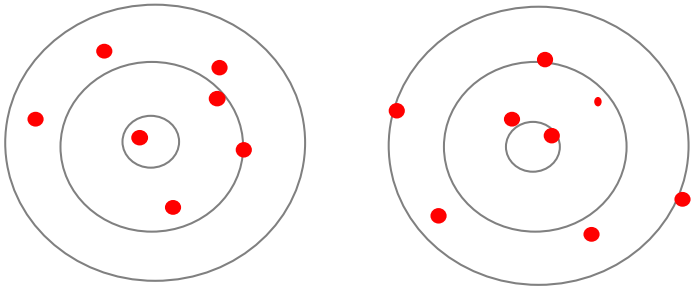



**Figure 1.2.** Relationship between bias and chance. Blood pressure measurements by intra-arterial cannula and sphygmomanometer.

# 1.1. DEFINITION

## Random error

- Low precision because of
  - Imprecise measuring
  - Too small groups



- Decreases with increasing group size

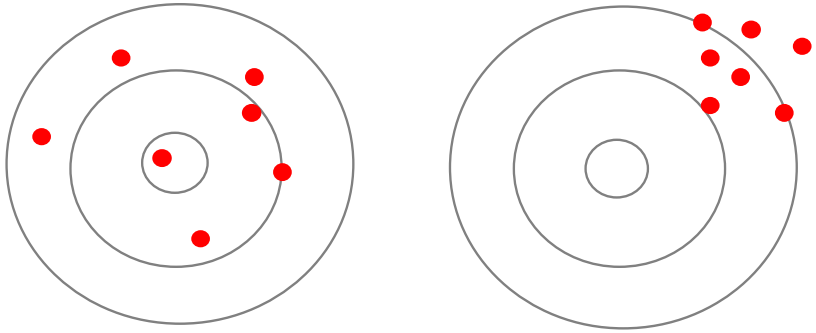
- Can be quantified by confidence interval

# 1.1. DEFINITION

## Systematic error

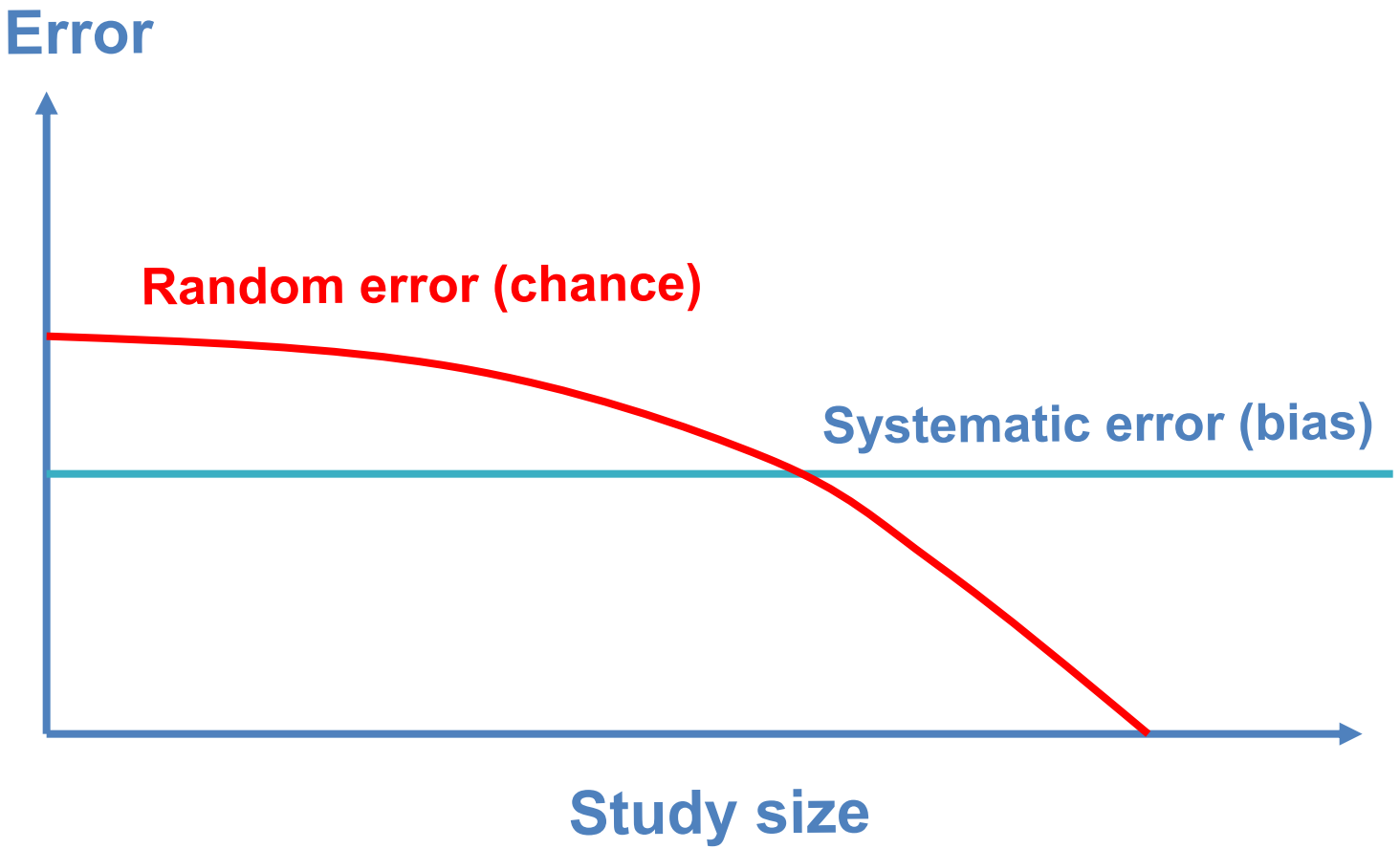
- Does not decrease with increasing sample size

- Selection bias
- Information bias
- Confounding



# 1.1. DEFINITION

## Errors in epidemiological studies



# 1. BIAS AND CONFOUNDING

1. Definition
2. Selection bias
3. Information bias
4. Confounding factor

# 1.2. SELECTION BIAS

**Definition + examples**

- Refers the method of collecting samples

- Examples:
  - Self-selection bias
  - Healthy worker effect
  - Non-response
  - Refusal

# 1. BIAS AND CONFOUNDING

1. Definition
2. Selection bias
3. Information bias
4. Confounding factor

# 1.3. INFORMATION BIAS

## Definition

- Error in the measurement of exposure or disease.

- Errors in the
  - procedures to measure exposure
  - procedures to diagnose disease

# 1.3. INFORMATION BIAS

**Examples of information bias**

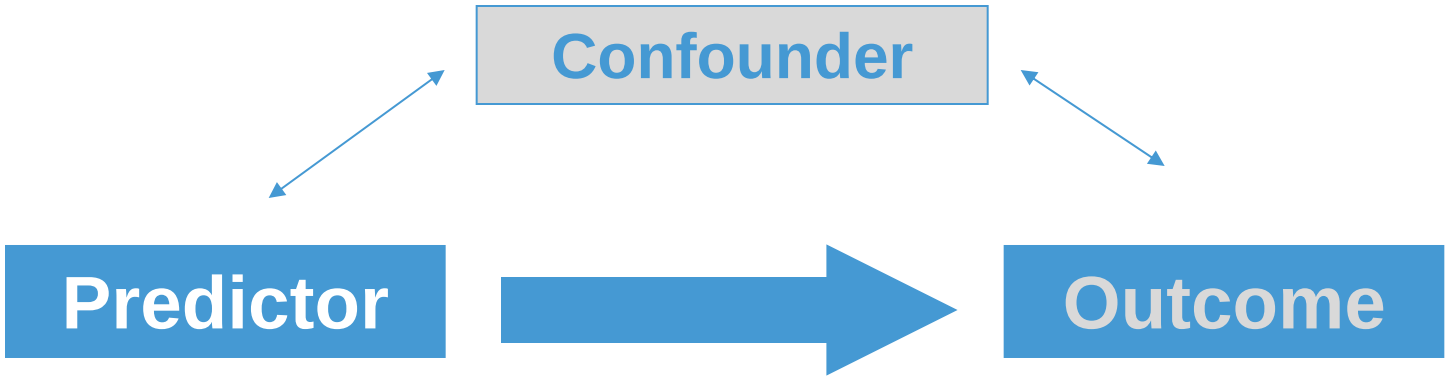
- Diagnostic bias
- Recall bias
- Researcher influence
- Wrong questionnaire

# 1. BIAS AND CONFOUNDING

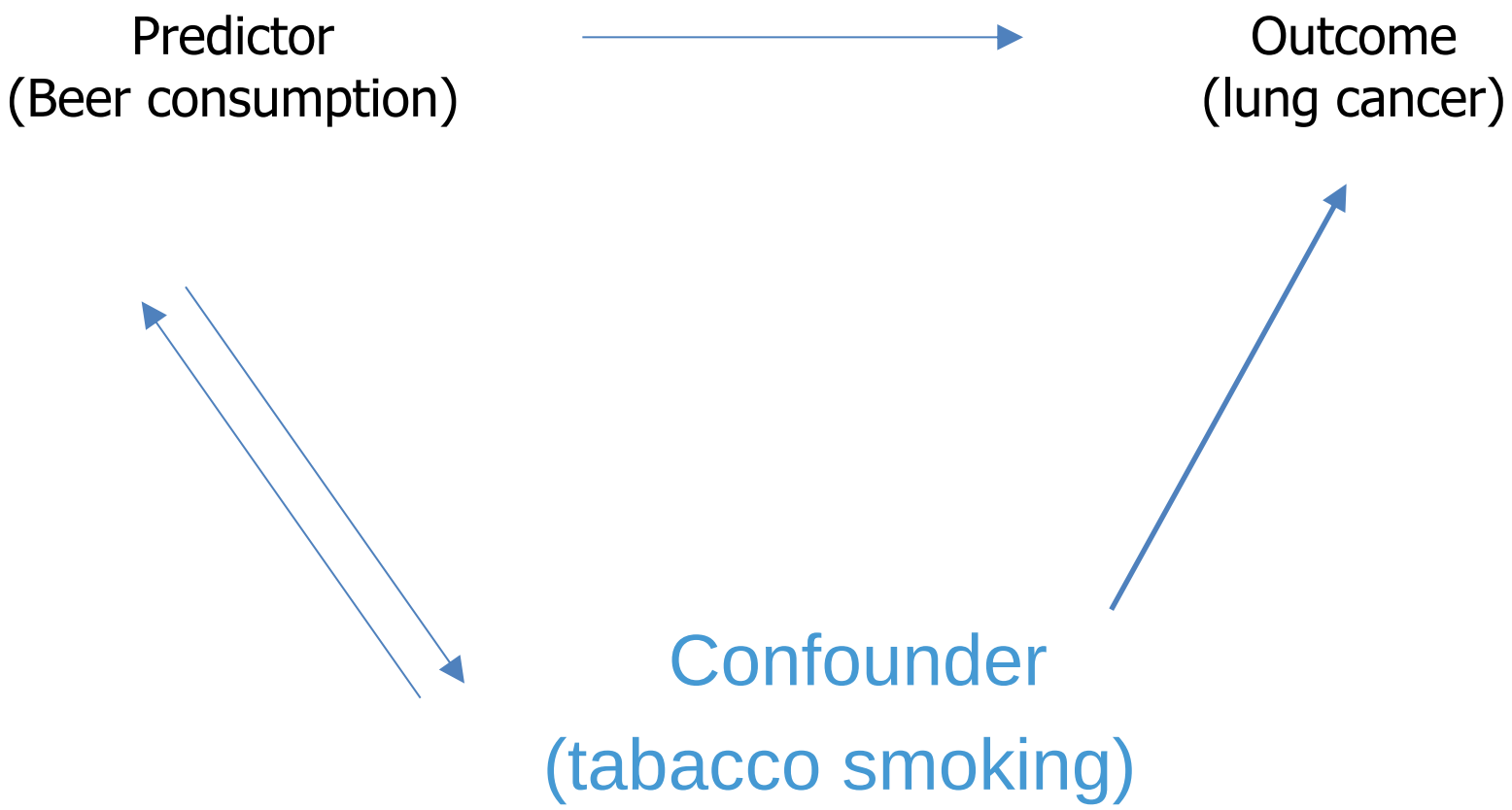
1. Definition
2. Selection bias
3. Information bias
4. Confounding factor

# 1.4. CONFOUNDING FACTOR

**Definition**



# **Exemple** : lung cancer and beer consumption



# EPIDEMIOLOGICAL STUDIES

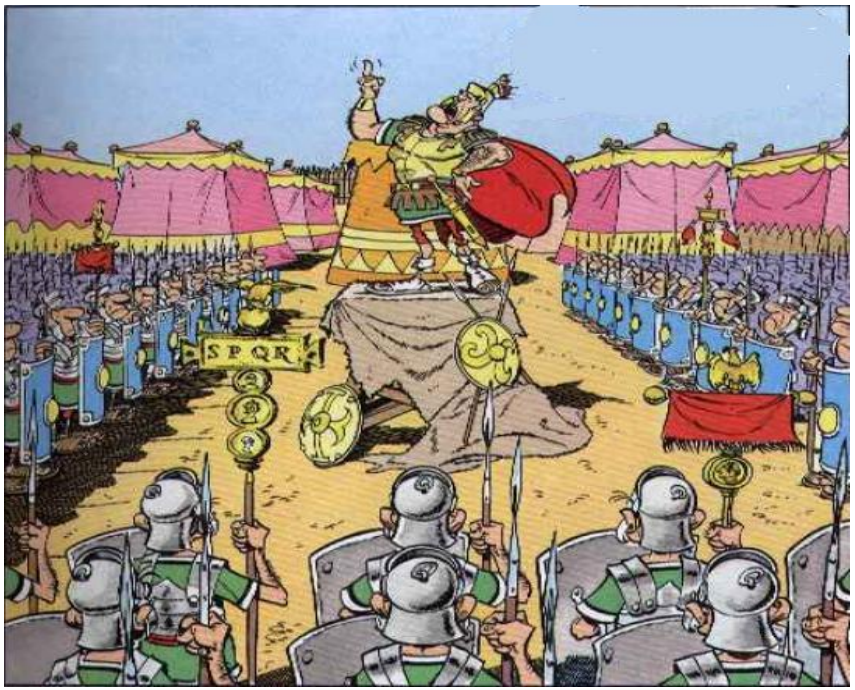
- Introduction
- Bias and confounding
- Cohort studies
- Case-control studies

# Cohort Studies

# COHORS-COHORTIS (LATIN) = "ENCLOSURE"

- Originally, the cohort was a sub-unit of a Roman legion, consisting of 480 legionaries including six centurions.

- …stayed together until all dead!



# DEFINITION

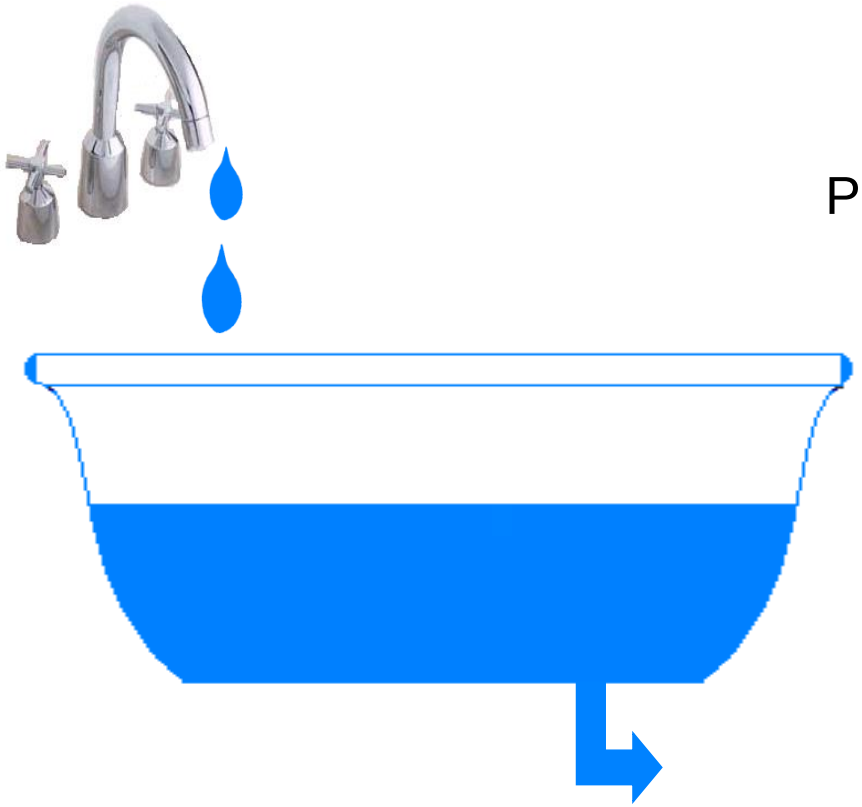
- Two or more groups of people, free of a disease or outcome at baseline, and differing according to exposure to a potential cause of disease, are followed forwards in time and compared for incidence of disease or death in each group.
- Other outcomes, too!

# INCIDENCE AND PREVALENCE

Aetiology - What's turning on the tap?

New **incident** cases

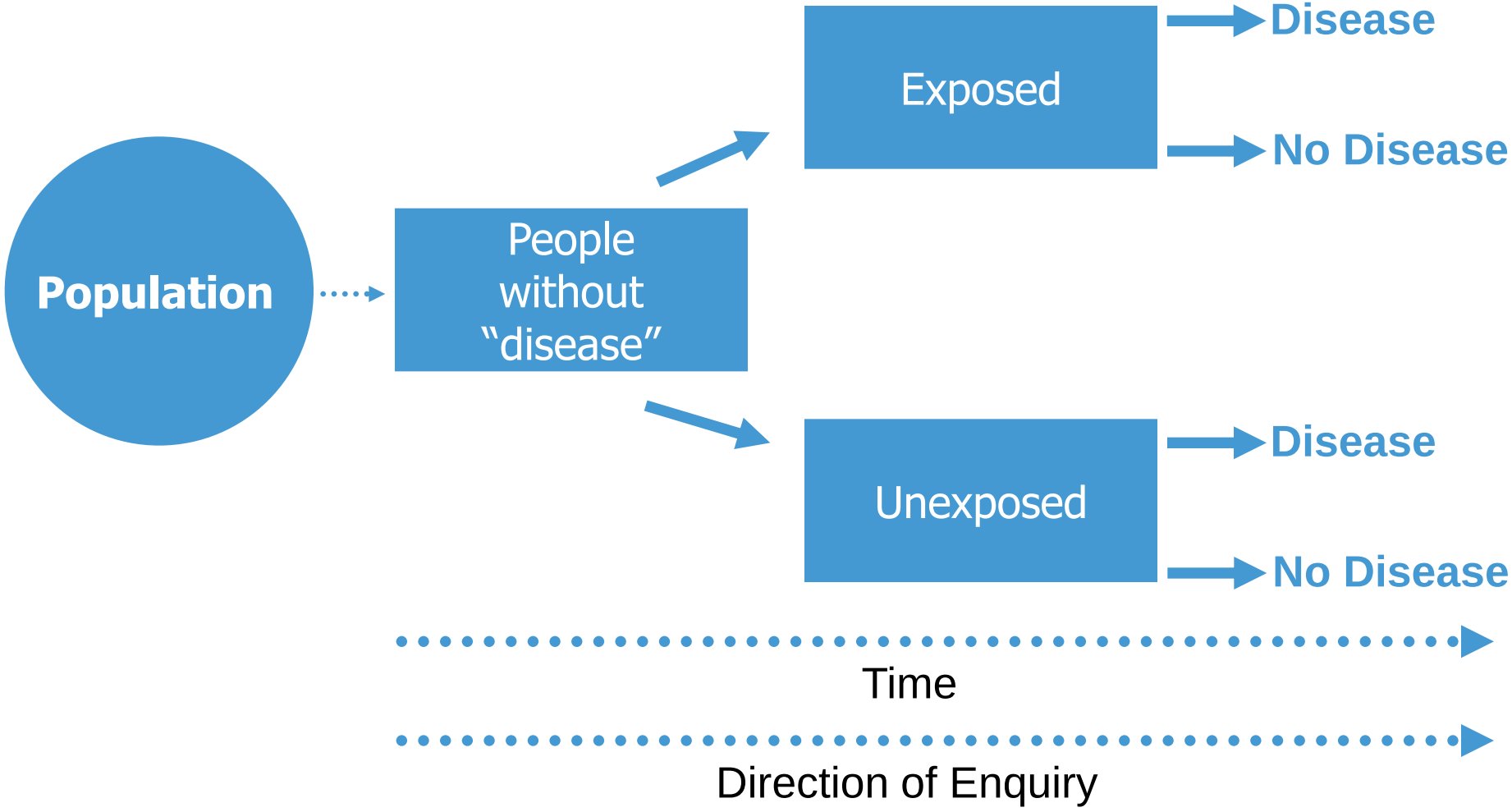
Pool of **prevalent** cases



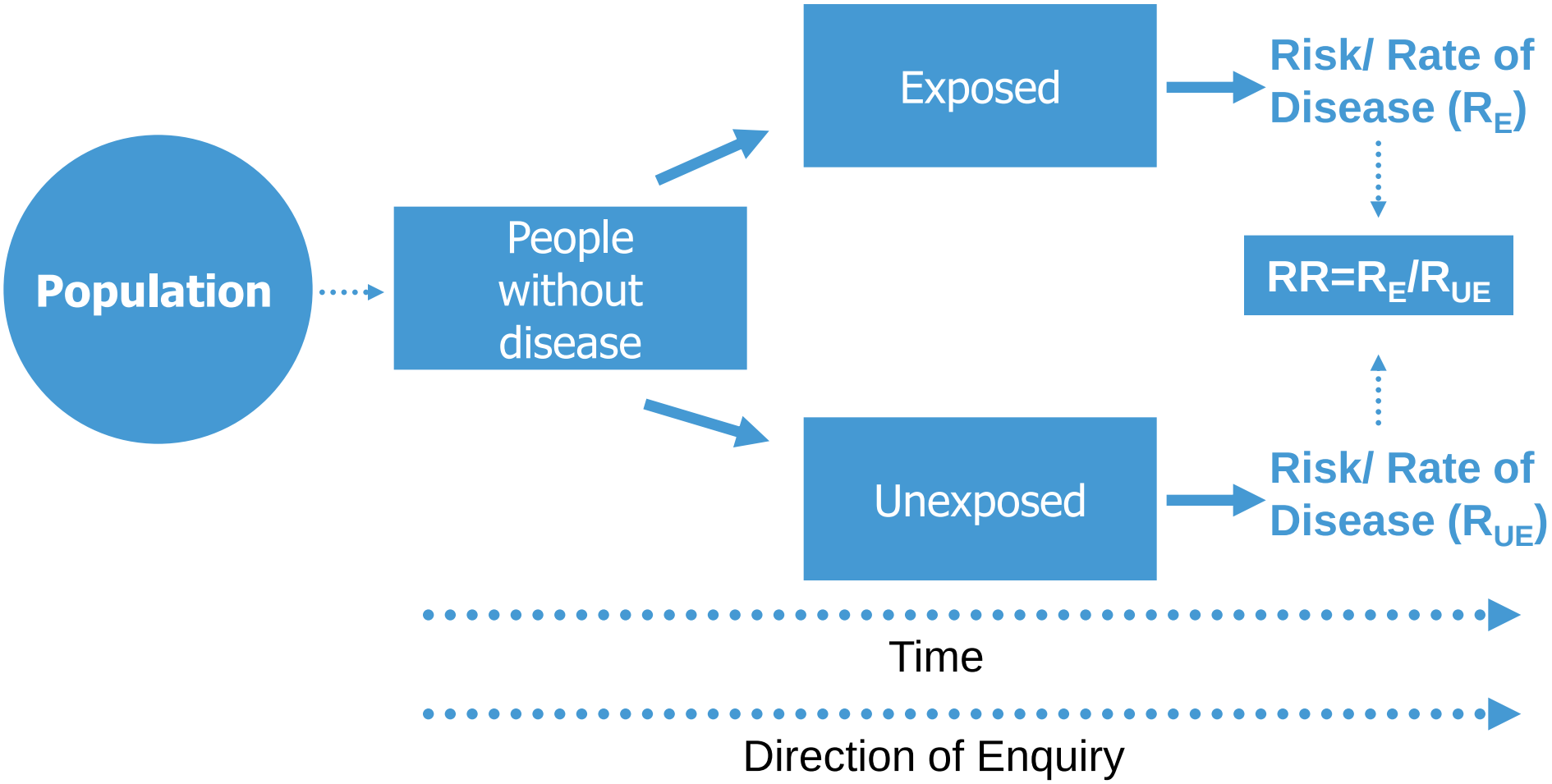
Prognosis

**Mortality**

# PROSPECTIVE OBSERVATIONAL COHORT



# PROSPECTIVE OBSERVATIONAL COHORT



# 2*2 TABLE

| | Disease | No Disease |
|---|---|---|
| Exposed | a | b |
| Unexposed | c | d |

- Observational Study with follow-up of incident cases over time
  - Incidence risk
  - Incidence rate (person years!)
- $R_E$=a/a+b
- $R_{UE}$=c/c+d
- RR=$R_E/R_{UE}$
- RR=(a/a+b)/(c/c+d)

# THIS IS SIMPLE: 2*2 TABLE

| | Disease | No Disease | Total |
|---|---|---|---|
| Exposed | a<br>2530 | b<br>7470 | a+b<br>10,000 |
| Unexposed | c<br>1265 | D<br>8735 | c+d<br>10,000 |

- Hypothetical cohort of 20,000 participants
- Incidence risk
  - Number of new cases of a disease in a given time period/ Number of disease free persons at beginning of that time period
  - 2530/10,000=0.2530 ($Risk_E$)
  - 1265/10,000=0.1265 ($Risk_{UE}$)
  - Risk Ratio=0.2530/0.1265=2

# THIS IS SIMPLE: 2*2 TABLE

| | Disease | No Disease | Total |
|---|---|---|---|
| Exposed | a<br>2530 | b<br>7470 | a+b<br>10,000<br>**78,000 person-years** |
| Unexposed | c<br>1265 | D<br>8735 | c+d<br>10,000<br>**92,503 person-years** |

- Incidence rate
  - Number of persons who have become cases in a given time period/ total person time at risk
  - 2530/78,000=0.0324 ($Rate_E$)
  - 1265/92,503=0.0137 ($Rate_{UE}$)
  - Rate Ratio=0.0324 /0.0137=2.36

# RISK/ RATE RATIOS

RR=1

- Disease rate among the exposed is the same as disease rate in the unexposed

RR>1

- Increased risk among the exposed

RR<1

- Decreased risk among the exposed, and the factor may be protective

# STRENGTHS OF COHORT STUDIES

- Temporal relationship
  - Cause (risk factor) must precede effect (disease)
  - New cases of disease (incidence)
  - Reduce biases e.g. due to reverse causality or recall bias
  - Test new hypotheses after study has started
- Rich phenotyping
  - Multiple disease outcomes (risks and benefits) of a given exposure
  - Multiple exposures on one disease outcome
  - Multiple measures over time
- Estimation of attributable risk and relative risk, and incidence rate

# WEAKNESSES OF COHORT STUDIES

- Duration, Size/ Power
- Costs (expensive compared to case control studies)
- Not useful for rare events
- Difficult to clinically characterise subphenotypes
- Phasic data collection: retrospective ascertainment
- Current practice/ exposure may change over time
- Biases, particularly loss to follow up (attrition)
- Confounding
- Multiple testing

# EPIDEMIOLOGICAL STUDIES

- Introduction
- Bias and confounding
- Cohort studies
- Case-control studies

# CASE-CONTROL AND CROSS-SECTIONAL STUDIES

# OUTLINE

- Case control studies
  - Definition
  - Design issues
  - Bias
  - Basic methods of analysis
- Cross-sectional studies
  - Definition

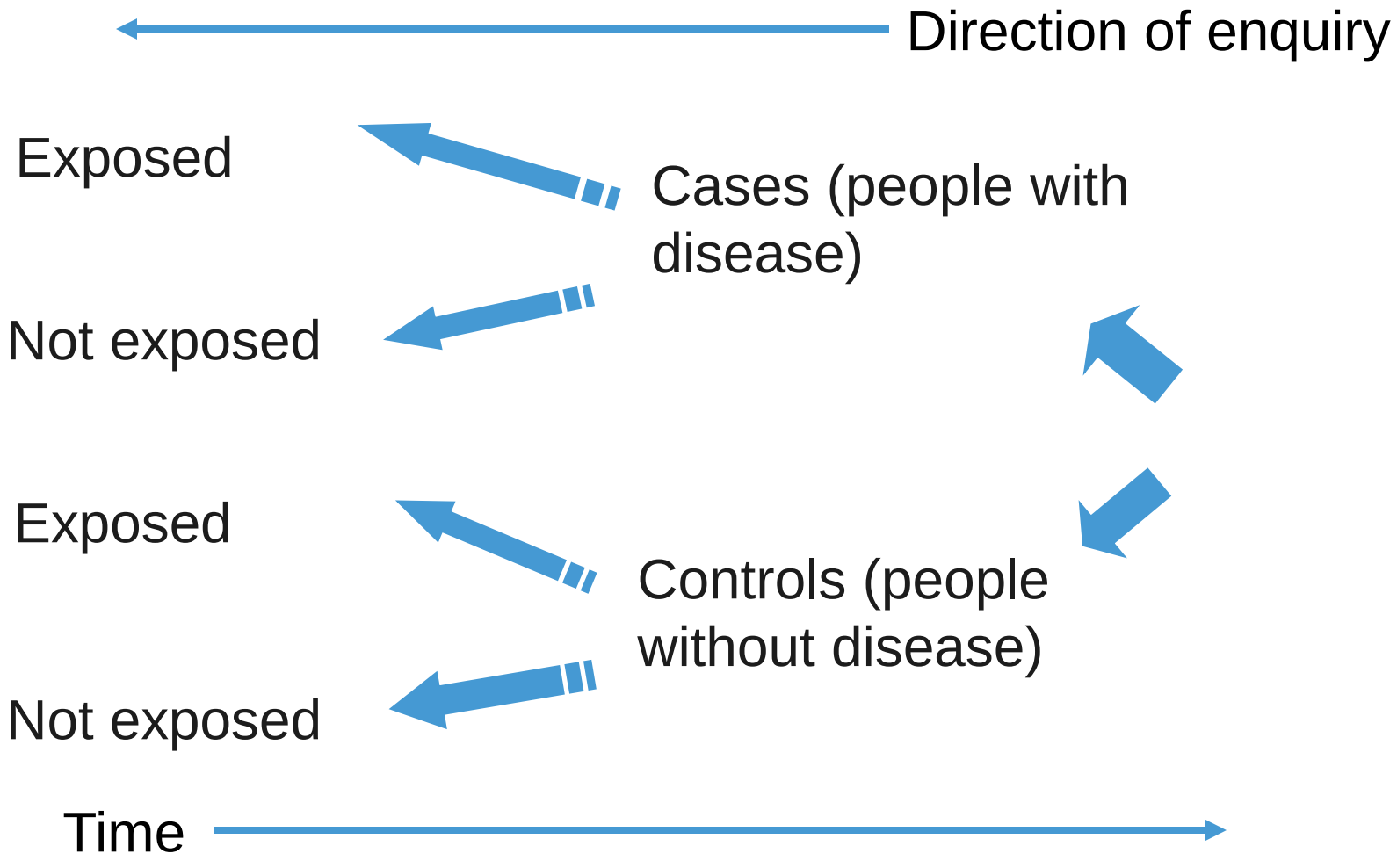
# WHAT IS A CASE-CONTROL STUDY?

- In a case-control study, two groups of people are considered:

- Cases - people with the disease of interest
- Controls - people without the disease

- Cases and controls are then compared with respect to their exposure to different risk factors of interest

# DESIGN OF A CASE-CONTROL STUDY



# ANALYSIS

In a simple unmatched study with exposed and unexposed groups the data can be presented as follows:

| | Cases | Controls |
|---|---|---|
| Exposed | a | b |
| Unexposed | c | d |

# ANALYSIS

- **Use odds ratio** as an estimate of the relative risk

# ODDS RATIO AS AN APPROXIMATION TO THE RELATIVE RISK

- The odds ratio **ad/bc** in a case-control study provides an approximation to the relative risk.

- This is the ratio of the odds of exposure in the cases

  **a/c**

- to the odds of exposure in the controls

  **b/d**

| | Cases | Controls |
|---|---|---|
| Exposed | a | b |
| Unexposed | c | d |

# GASTRIC CANCER AND CHILLI PEPPER CONSUMPTION

| | | Cases | Controls |
|---|---|---|---|
| Chilli pepper | Yes | 211 | 607 |
| | No | 9 | 145 |

Odds ratio (OR) $= \frac{211 \times 145}{607 \times 9} = 5.6$

95%CI: 2.7 to 12.0

# DIET AND ENDOMETRIAL CANCER

| | | Cases | Controls |
|---|---|---|---|
| Dietary fibers | Highest third | 33 | 81 |
| | Lowest third | 46 | 78 |

Odds ratio (OR) $= \frac{33 \times 78}{81 \times 46} = 0.69$

95% confidence interval is 0.57 to 2.83

# CONTROL SELECTION

- Usually the hardest part of a case-control study
- Controls should be drawn from the population at risk of becoming cases
- Can consider controls as being a sample drawn from a large cohort study from which the cases were identified
- It should be possible to ascertain exposure in a similar way as for the cases
- Aim to eliminate bias

# SOME TYPES OF CONTROLS:

- Hospital controls
- Community controls:
  - GP registers
  - Electoral registers
  - Birth registers
  - Neighbourhood

# TO MATCH OR NOT TO MATCH?

# MATCHING

- Controls can be matched individually to each case on a variety of confounding factors
- 1:1 or 1:n matching. 1:1 matching often called ‘matched pairs’
- Rarely sensible to have more than 4 controls per case
- Common matching variables are age and sex

# CASE-CONTROL STUDY

Advantages:

- rare outcomes
- relatively cheap
- fairly quick

Disadvantages:

- exposure measure liable to bias
- no good for rare exposures
- selection of controls vital

# CROSS-SECTIONAL STUDIES


## Advantages:

- cheap & quick

## Disadvantages:

- no time element
- not good for rare outcome

**or**

- exposures

# CROSS-SECTIONAL STUDIES

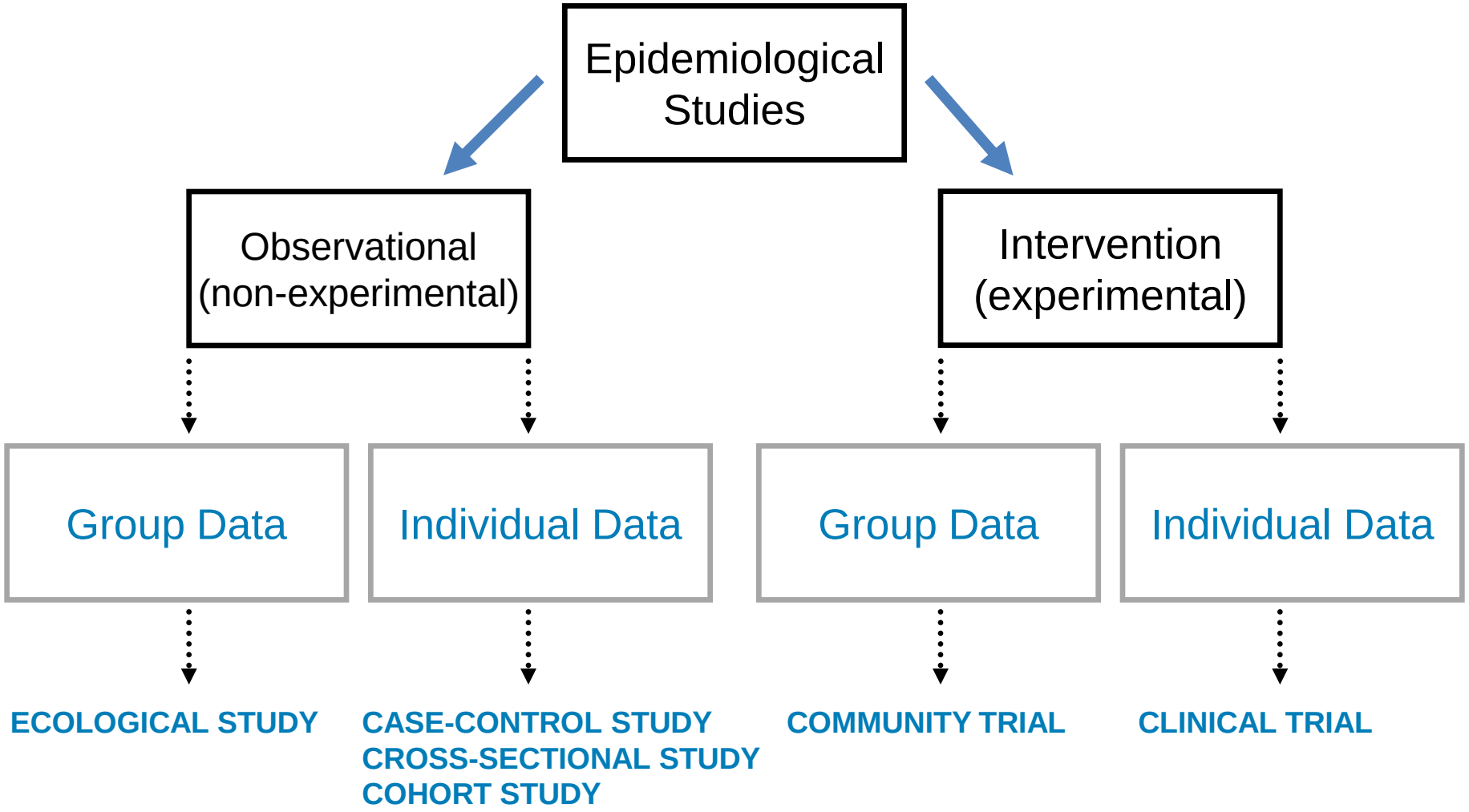
- Any general survey is a cross-sectional study
- Sometimes known as prevalence studies
- Data collected from a sample of the population, such as everyone in an occupational group, a specific village, or a random sample from a geographical area
- Often data collected by questionnaire
- Useful for ascertaining prevalence of disease and exposures for planning purposes, but can be used to search for aetiological factors

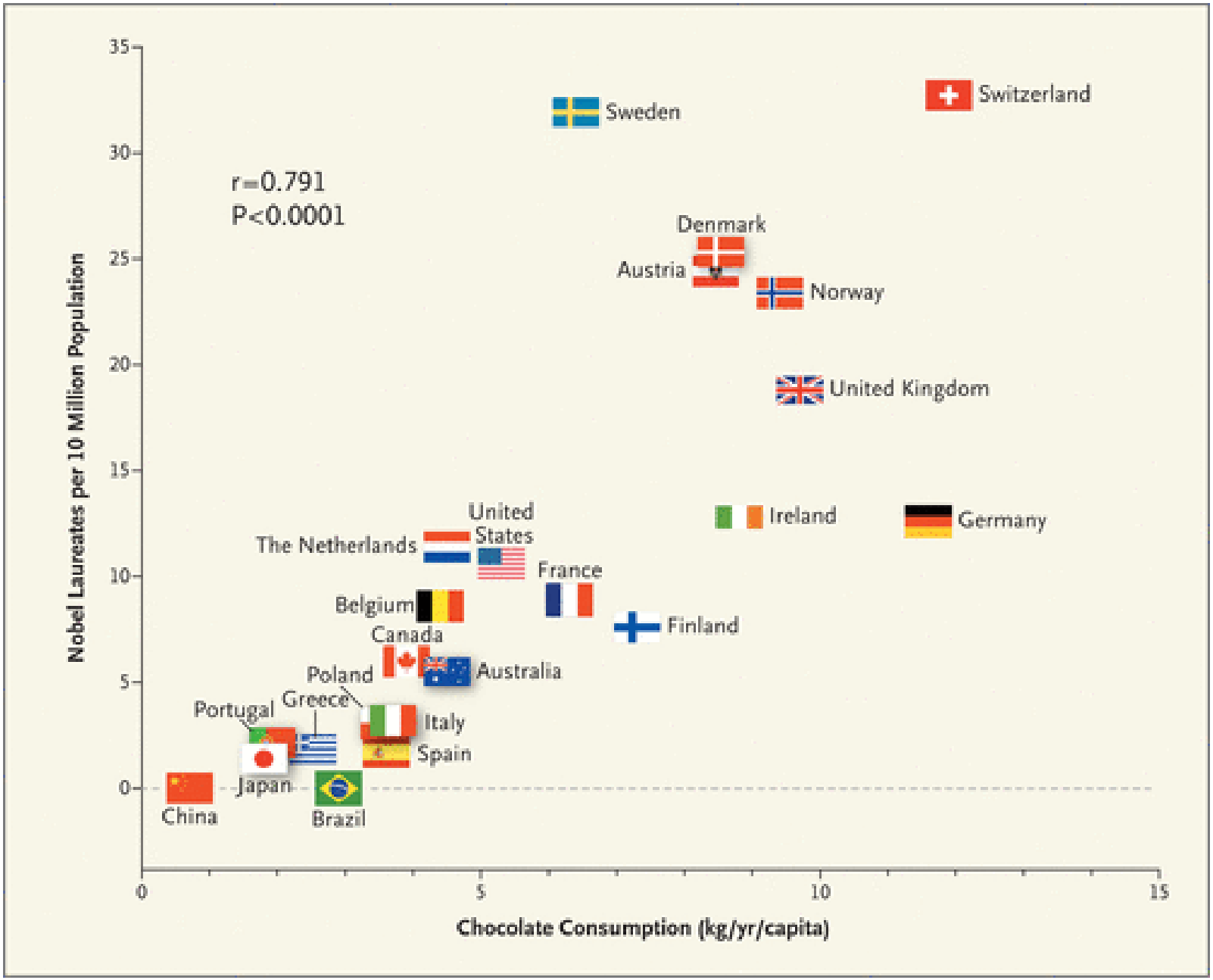
# CONCLUSION

"The data are the object of inquiry rather than the character and intelligence of those who generate it……individuals of highest intelligence can generate flawed information and those of limited talent can stumble into trustworthy findings" (Savitz – epi evidence)

# Just for fun...

# STUDY DESIGN OVERVIEW




Nobel Laureates per 10 Million Population
Chocolate Consumption (kg/yr/capita)
r=0.791
P<0.0001
Sweden
Switzerland
Denmark
Austria
Norway
United Kingdom
Ireland
Germany
United States
The Netherlands
France
Belgium
Finland
Canada
Australia
Poland
Greece
Portugal
Italy
Spain
Japan
China
Brazil
0
5
10
15
20
25
30
35

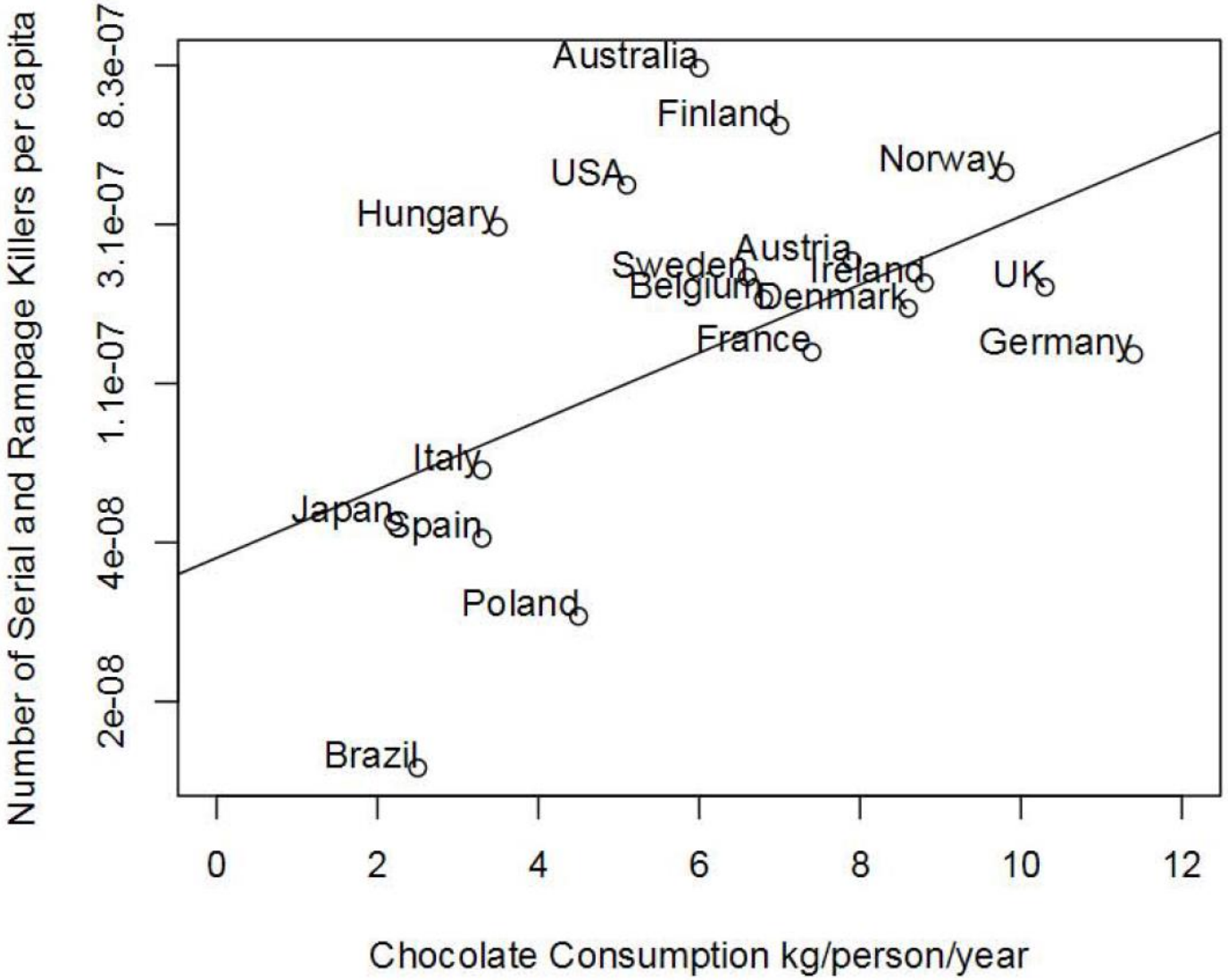



**Figure 1**: Correlation between countries' annual per capita chocolate consumption and the serial and rampage killers per capita since 1900.