



# **ERROR in epidemiological studies**

**Taught by Dr. Pratap Singhasivanon**

Wrap up by Dr. Chollasap Sharma & Dr. Ranida Poksiri

# Objectives

- The concept of systematic errors.
- Types of systematic error.
- Misclassification
- Confounding variables
- Methods to control confounding



# Coffee and Cancer of the Pancreas

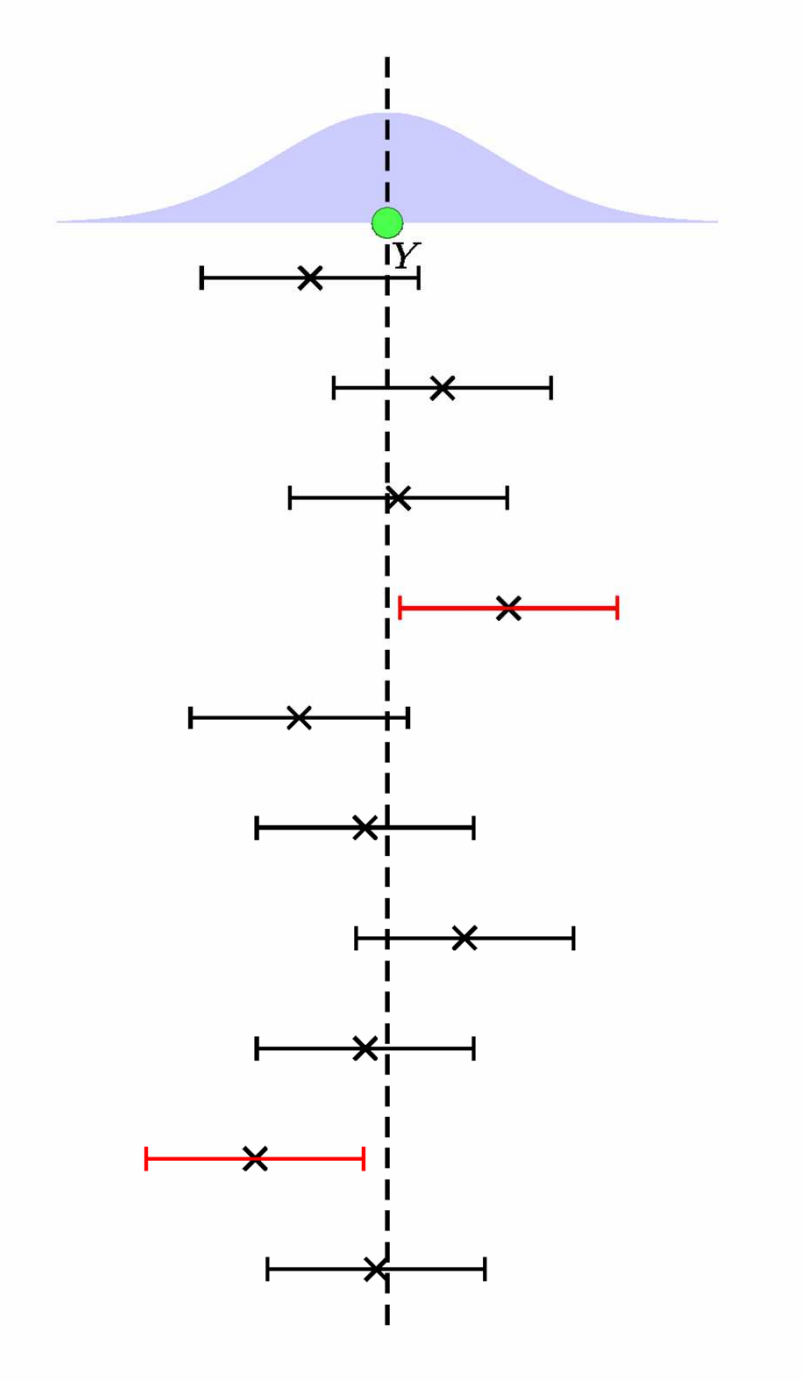
Brian MacMahon, M.D., Stella Yen, M.D., Dimitrios Trichopoulos, M.D., Kenneth Warren, M.D., and George Nardi, M.D.

N Engl J Med 1981; 304:630-633 March 1981

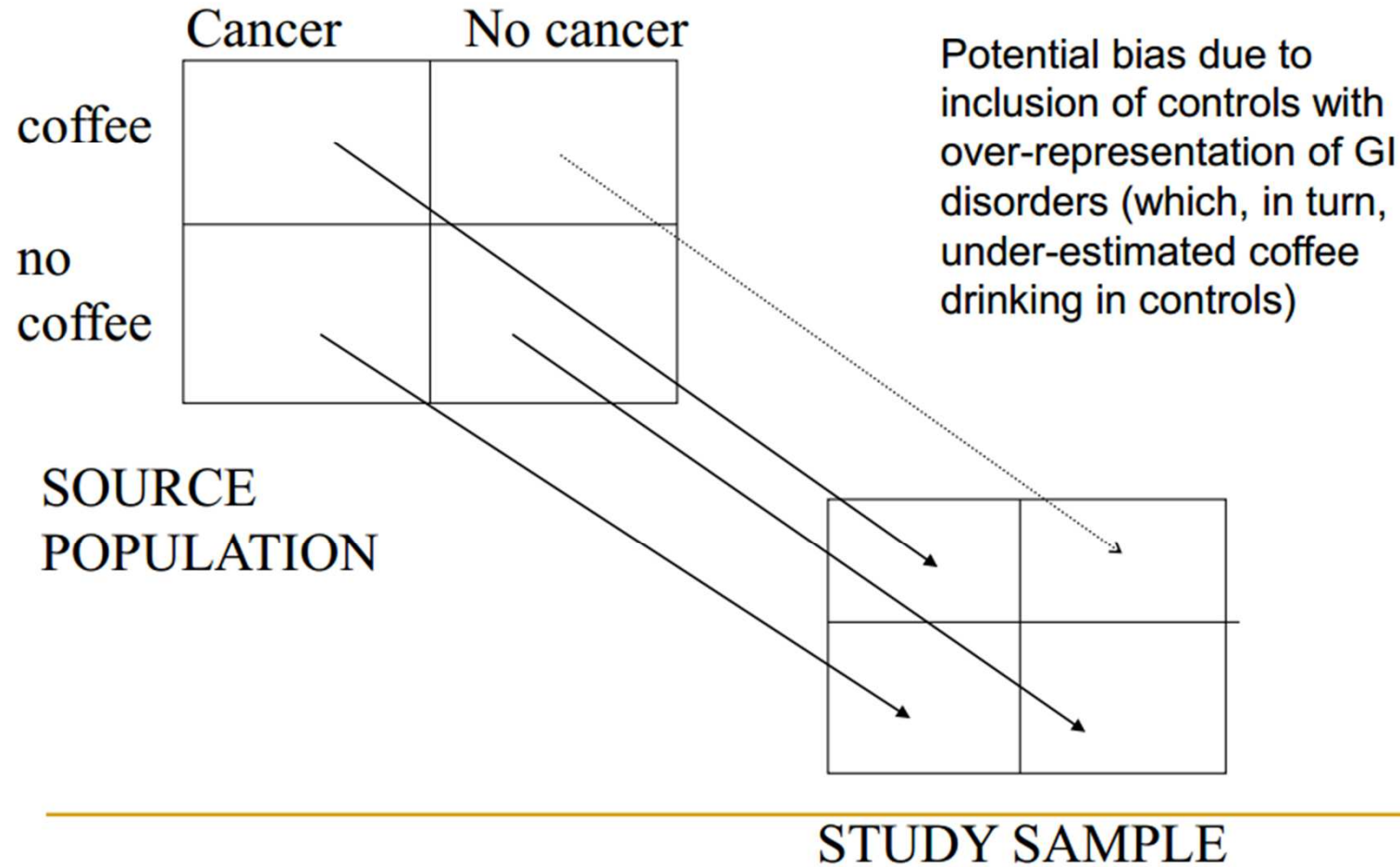
Drinking coffee  CA Pancrease

**OR = 2.7**

With  $\geq 3$  cups / day = 2.7 ;  
95% C.I (1.6 to 4.7).



# Case-control Study of Coffee and Pancreatic Cancer: Selection Bias



## Coffee and cancer of the pancreas: Use of population-based controls

- Gold et al. *Cancer* 1985

Case Control

Coffee: $\geq 1$ cup day	84	82
No coffee	10	14

$$OR = (84/10) / (82/14) = 1.4 \text{ (95\% CI, 0.55 - 3.8)}$$

So, when population-based controls were used, there was no strong association between coffee and pancreatic cancer

# Goal of epidemiological study

- Accurate measurement of **factors & outcomes**



# Association

- A 'statistical dependence between  $\geq 2$  events, characteristics, or other variables'.

Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006



# Explanation for the observed difference

1. **Chance (Random error)**
2. **Bias (Systematic error)**
  - Selection
  - Information
  - Confounding
3. **Effect of exposure**



# FRAMEWORK FOR THE INTERPRETATION

## IS THERE A VALID STATISTICAL ASSOCIATION?

Is the association likely to be due **chance**?

Is the association likely to be due **bias**?

Is the association likely to be due **confounding**?

## CAN THIS VALID STATISTICAL ASSOCIATION BE JUDGED AS CAUSE AND EFFECT?

**ERROR** = **SYSTEMATIC ERROR** + **RANDOM ERROR**

**False result**  
obtained in a  
study

**BIAS**

Due to **factors** that  
inherent in the  
**design,**  
**measurement**  
**& analysis**

(RANDOM VARIABILITY)

**Result obtained in sample**  
**differs that would be**  
**obtained if the entire**  
**population were studies**

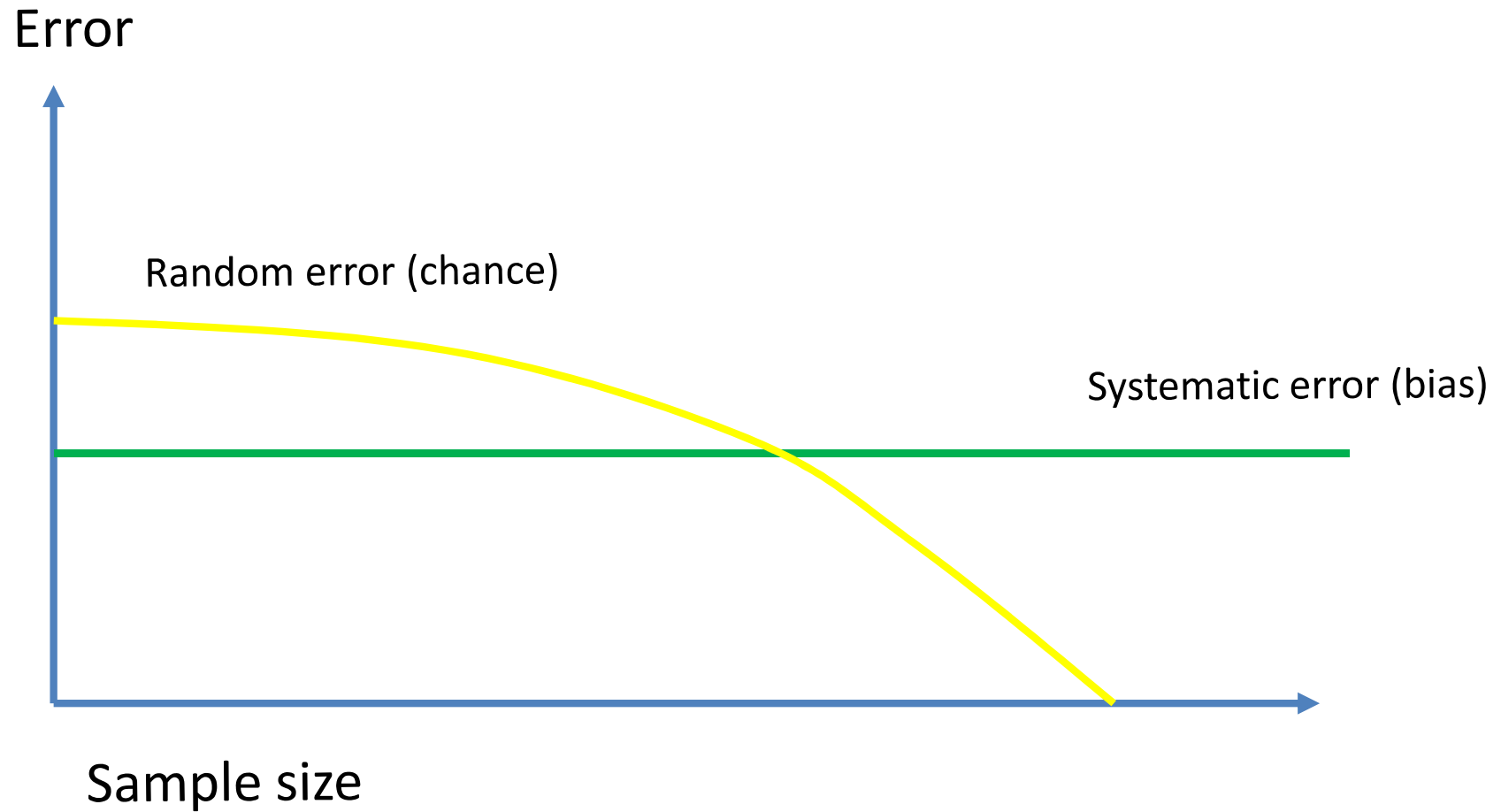
# RANDOM ERROR

Refers to **fluctuations** around a true value because of **Sampling variability**

# SYSTEMATIC ERROR

Any difference between the **true value** and that actually obtained i.e the result of **all causes** other than Sampling variability.

# Errors in epidemiological studies



# Bias

- Occurs when an estimated association (RR, OR, etc.) **deviates** from the true **measure of association**
- **Bias** may be introduced

# Classifying types of bias

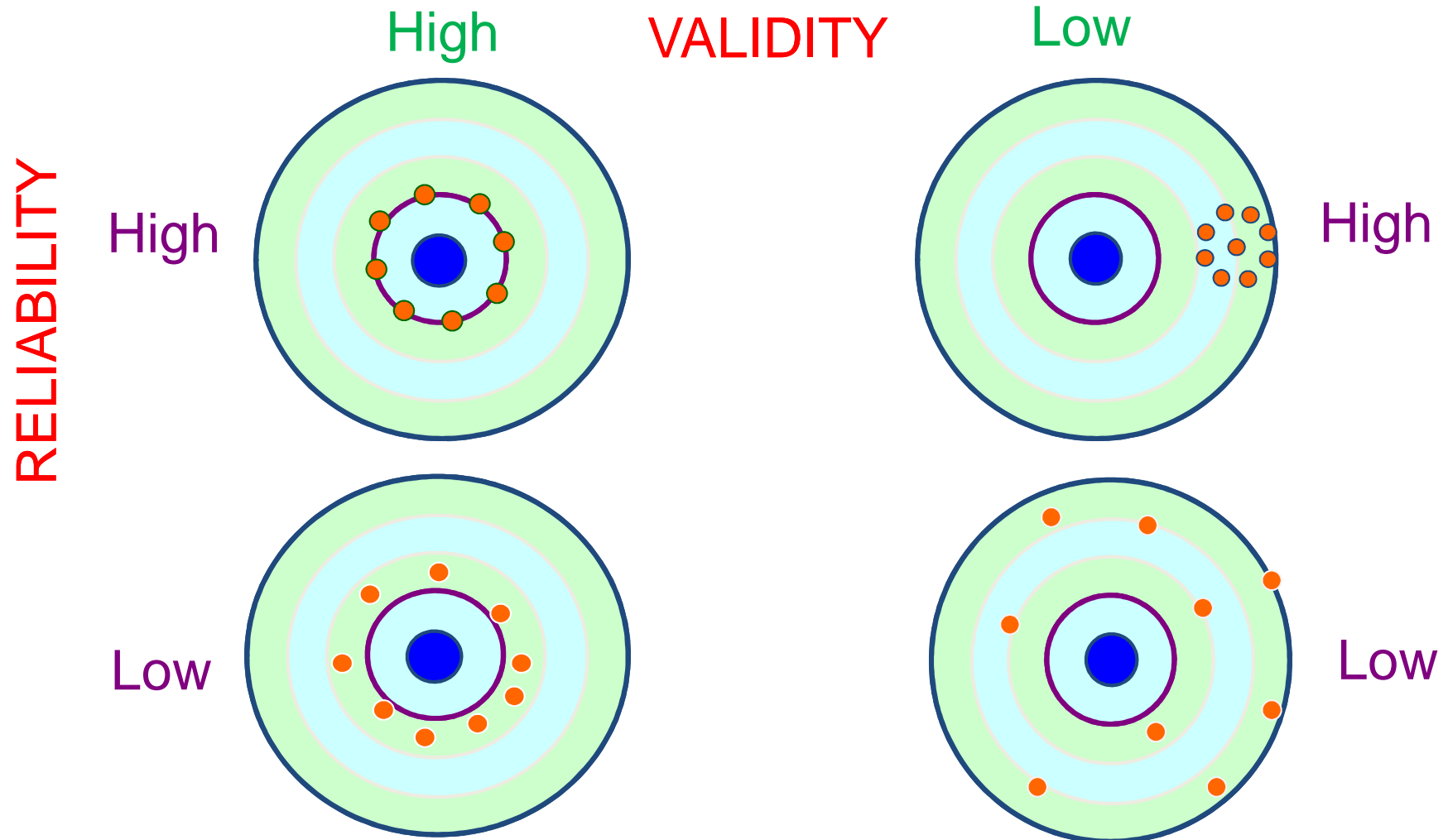
- **Selection bias** – differential access to the study population
- **Information bias** – inaccuracy in measurement or classification
- **Confounding bias** – unfair comparison



## VALIDITY :

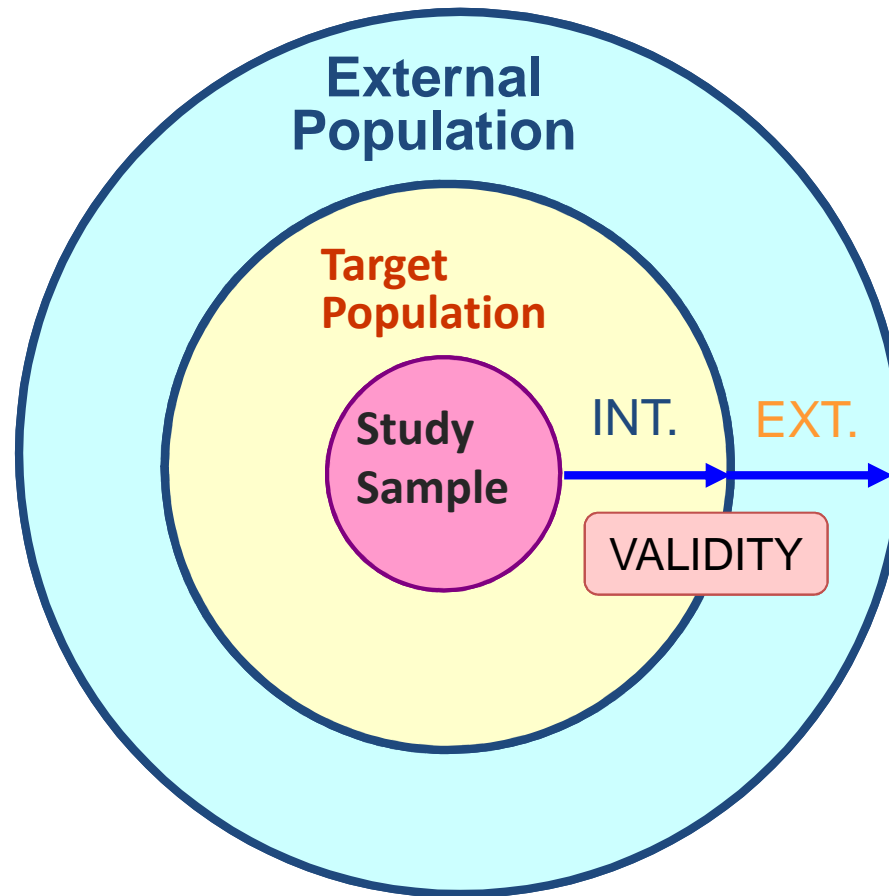
A study is valid if its results corresponds to the **truth**, **no systematic error** or should be as small as possible

# Different combinations of high and low reliability and validity



- ***Internal validity***: whether the study provides an unbiased estimate of **what it claims** to estimate
- ***External validity***: whether the results from the study can be generalized to some **other** population

# Internal and External Validity



# Selection Bias

- Distortions that arise from
  - Procedures used to select subjects
  - Factors that influence study participation
  - Factors that influence participant attrition

## Example:

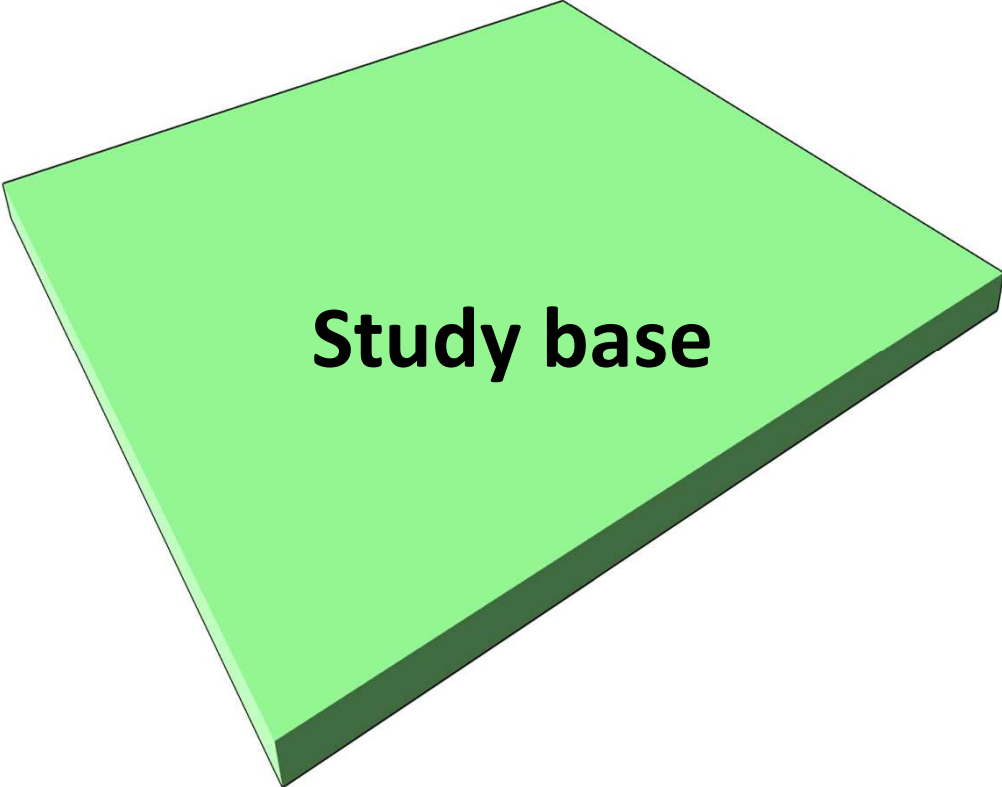
- If cases & controls or exposed & non-exposed individuals were selected in such a way that an association is observed even though exposure & disease are not associated
- May result from withdrawal or losses to F/U of study subjects

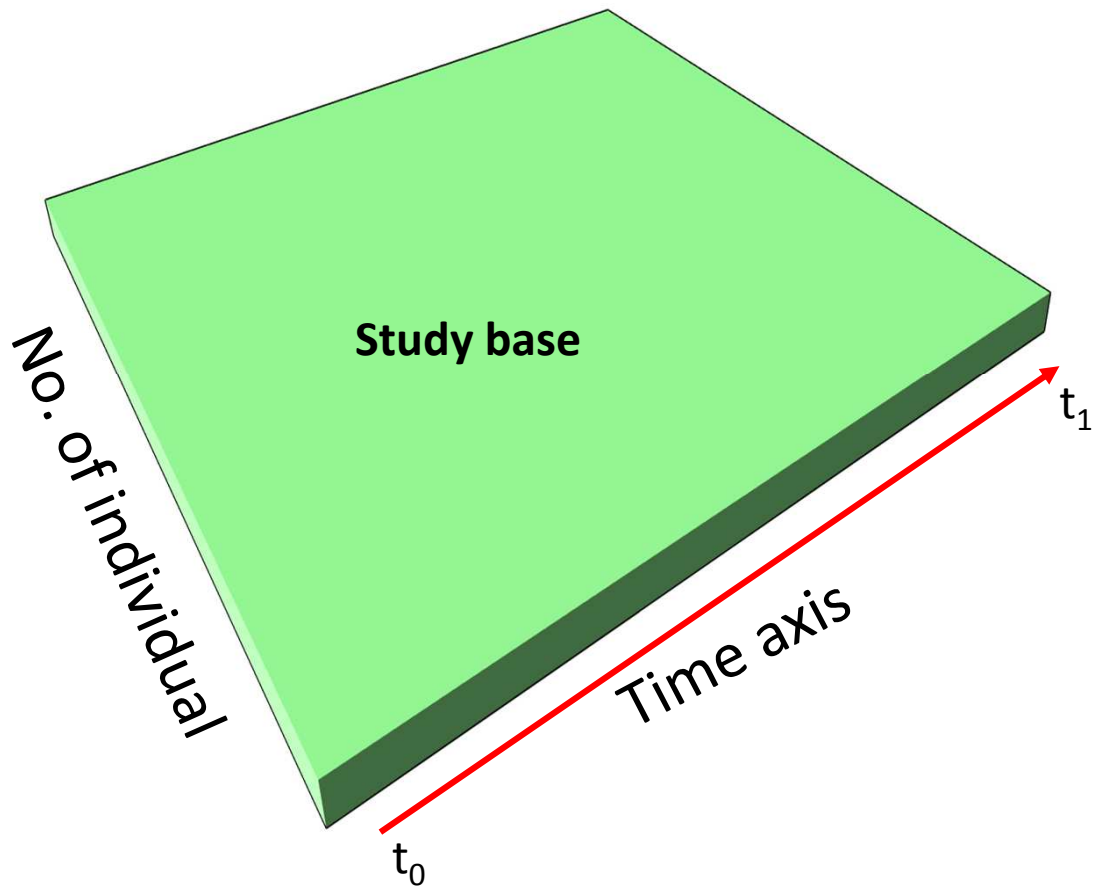
# Case-Control Study

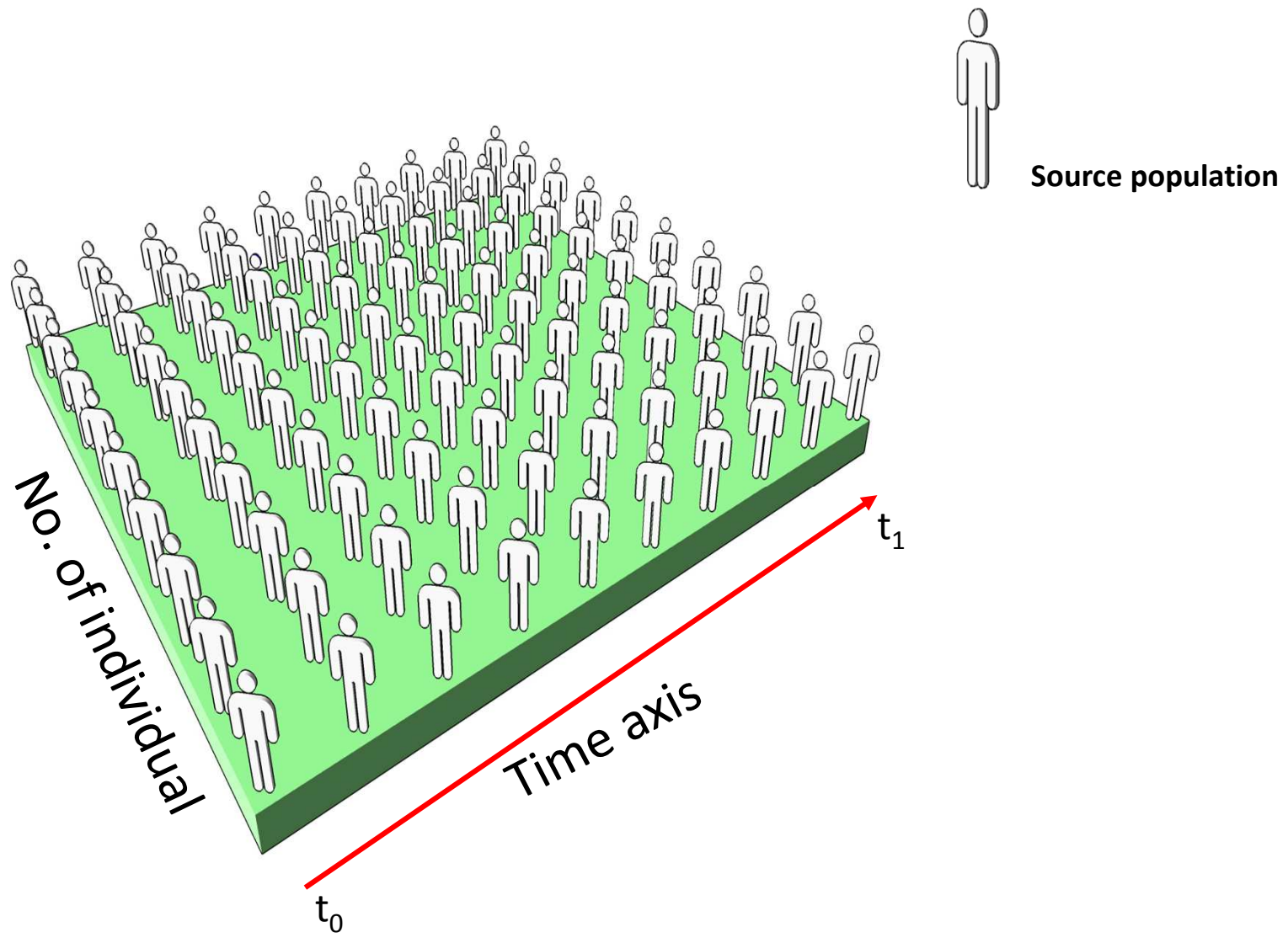
Case-control studies  
are prone to  
**selection bias**  
attributable to  
**flawed sampling of**  
**base populations.**

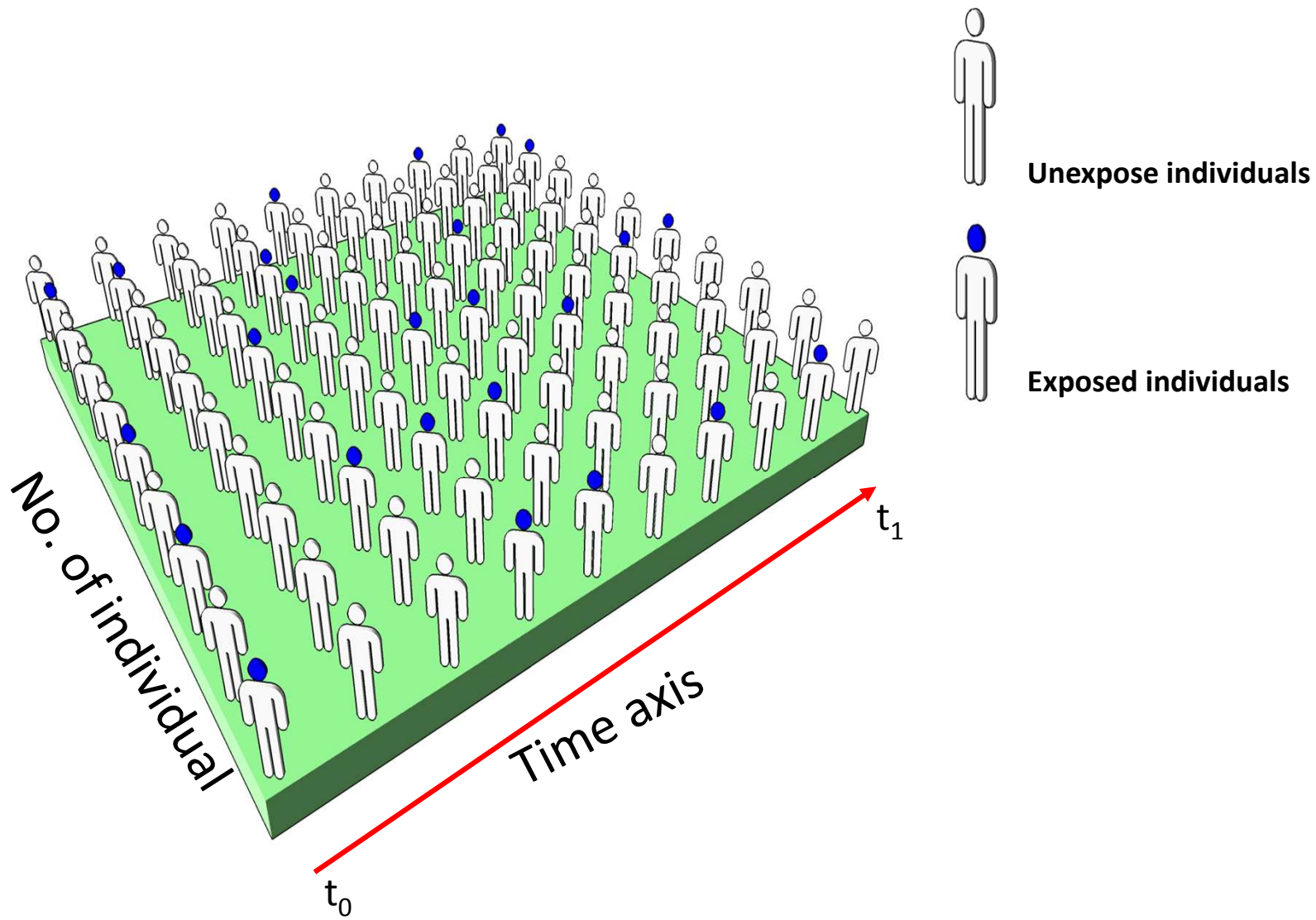


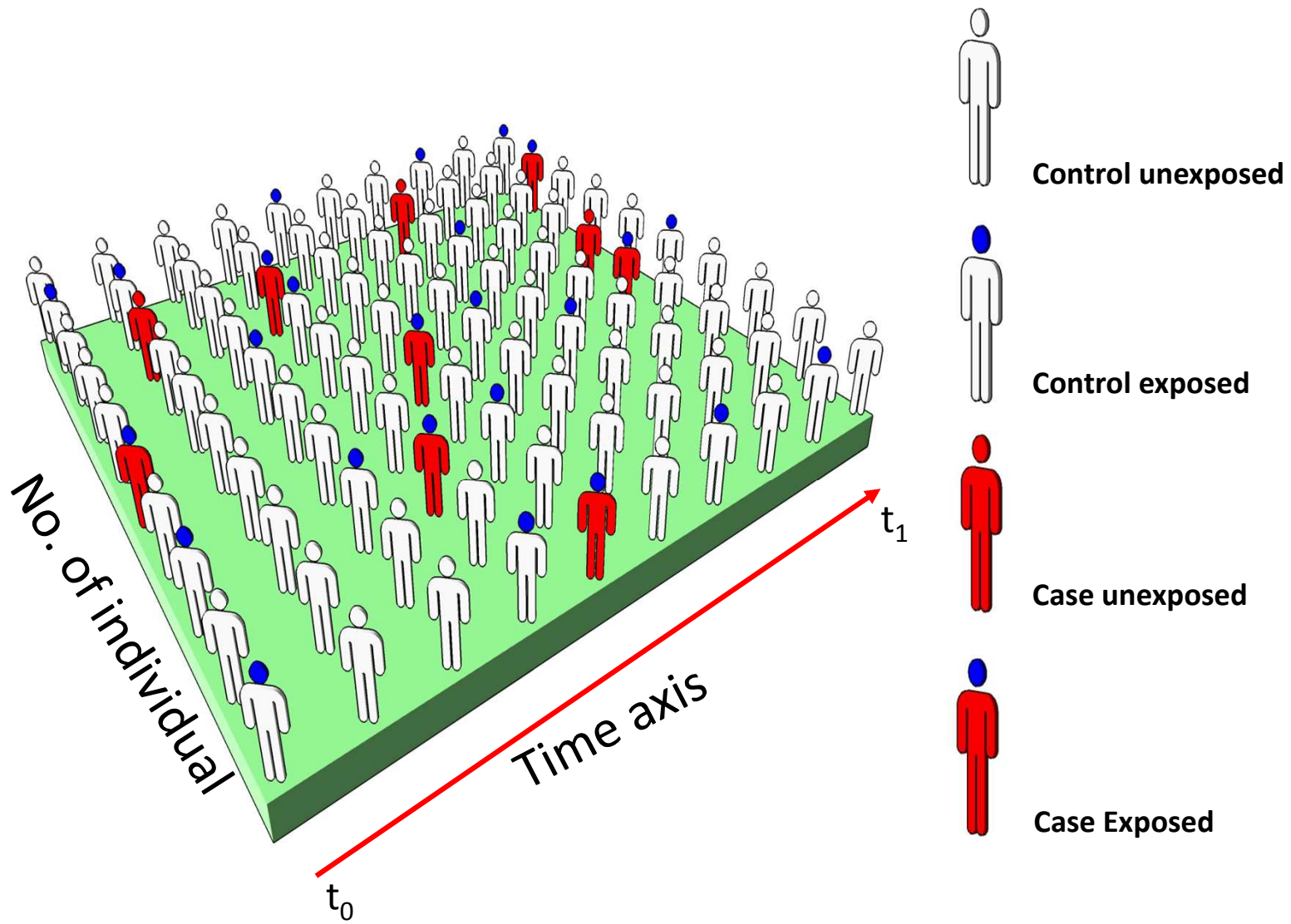
**“Cases and Controls should  
be representative of the same  
base experience”**



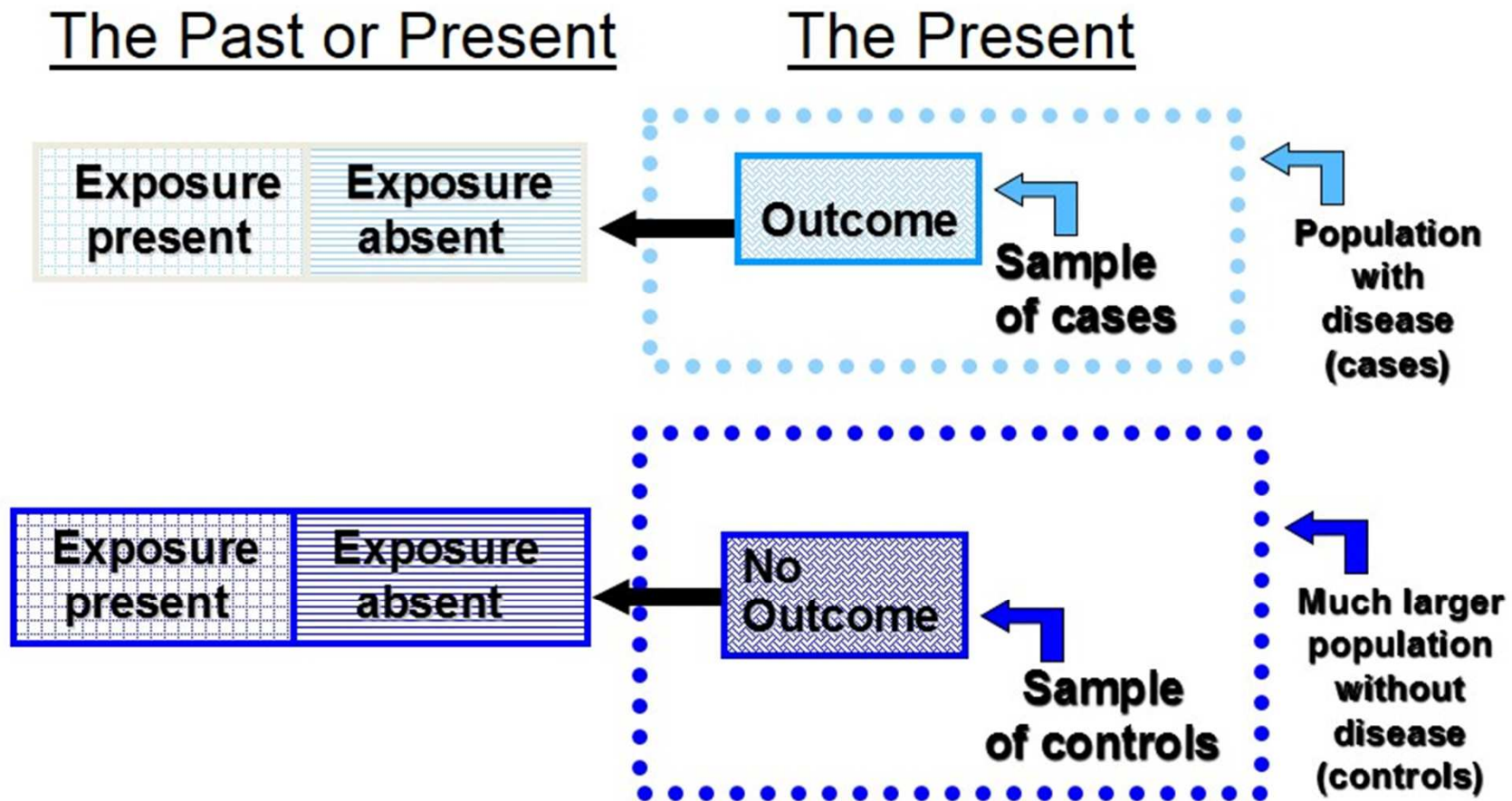








# Case-Control Design



The identification of the appropriate

**study base** (source population) from which to

select controls is the primary challenge in the design  
of case-control studies



# Selection Bias

- Can result in **over- or under- estimation** of the true magnitude of the relationship between an **exposure** and an **outcome**
- May **reduce** the validity of the study

# Selection Bias

- To avoid it, ensure that:
  - Subjects are representative of **target population**
  - Study and comparison groups are **similar** except for variables being investigated
  - Subject **losses** are kept to a minimum