CASE CONTROL STUDY

Case-control study

Exposure **Disease (+)**

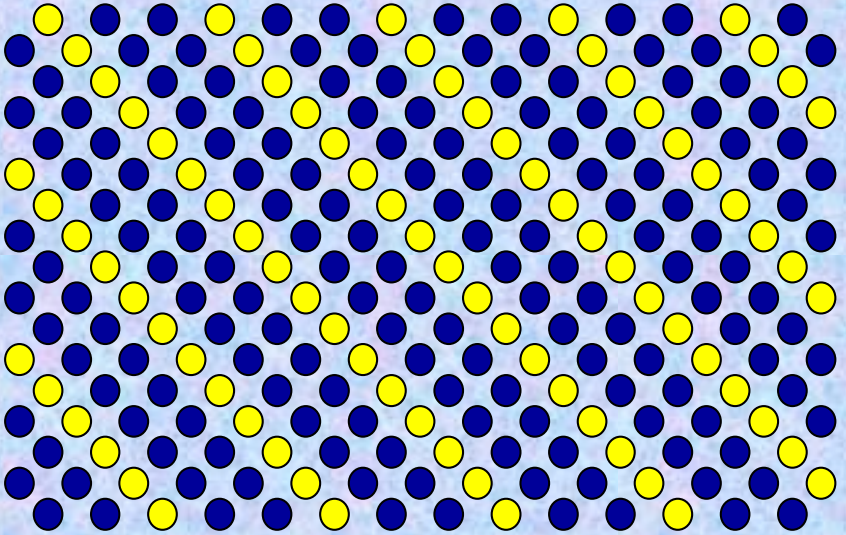
? ----------

Exposure **Disease (-)**

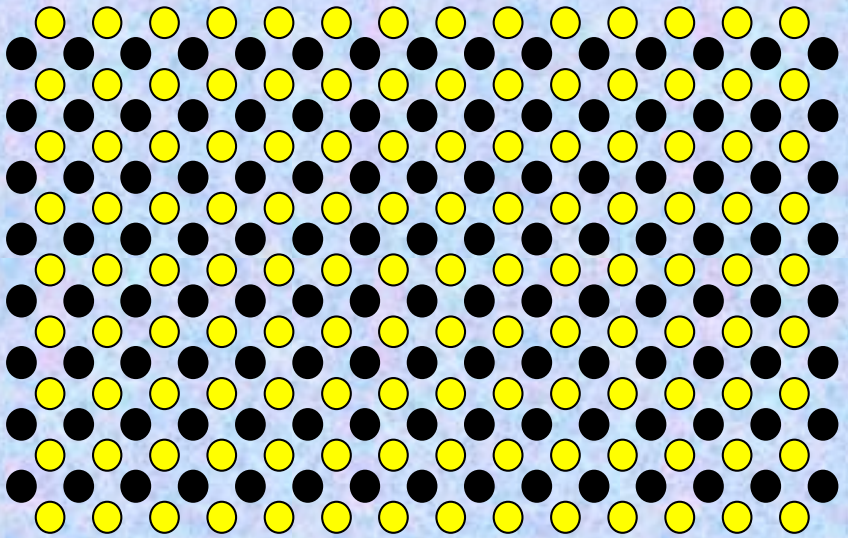
? ----------

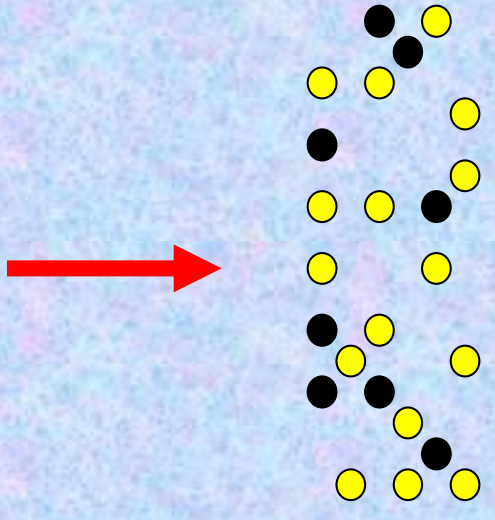
Investigator

Source population



- Exposed
- Unexposed

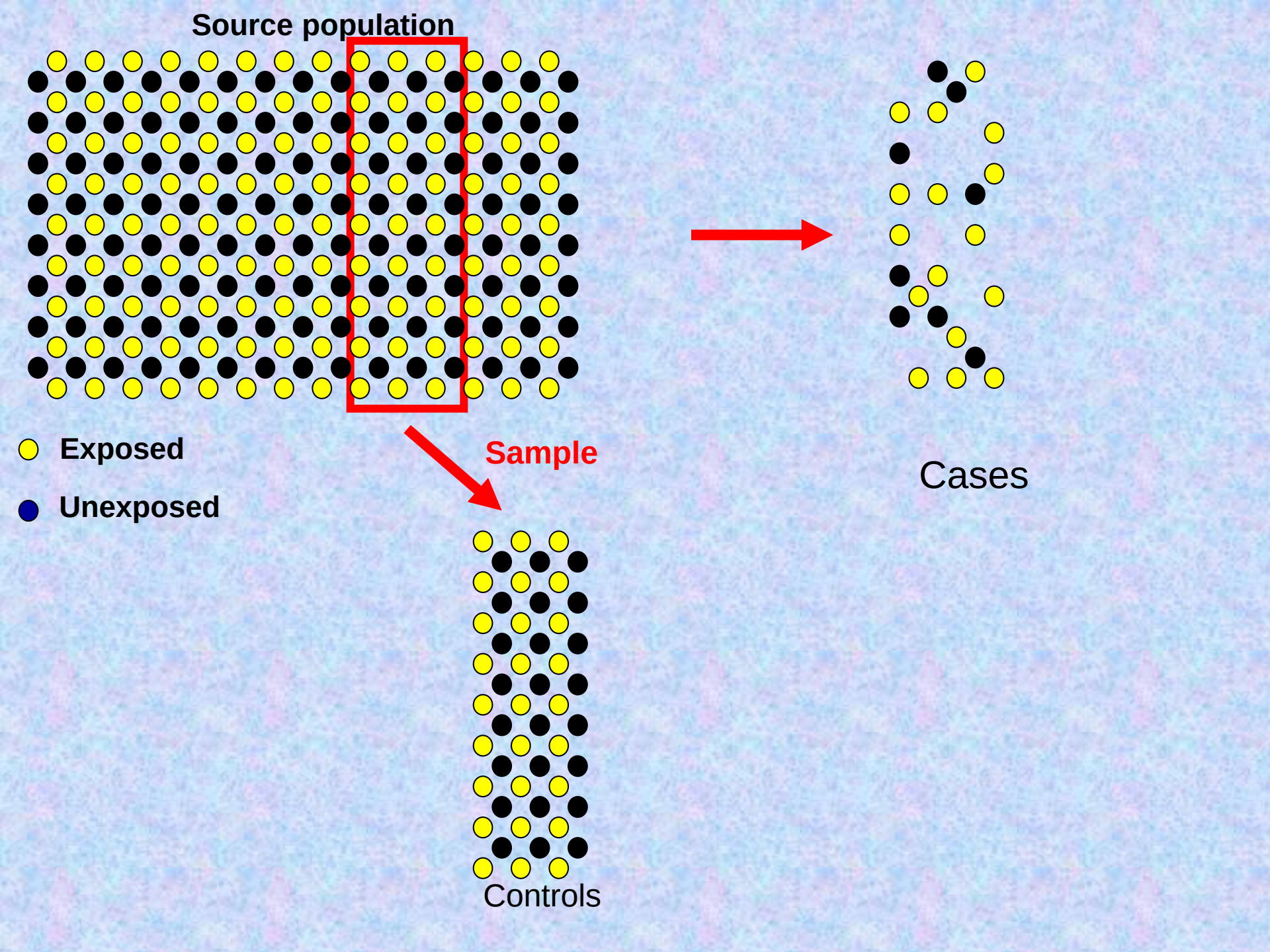


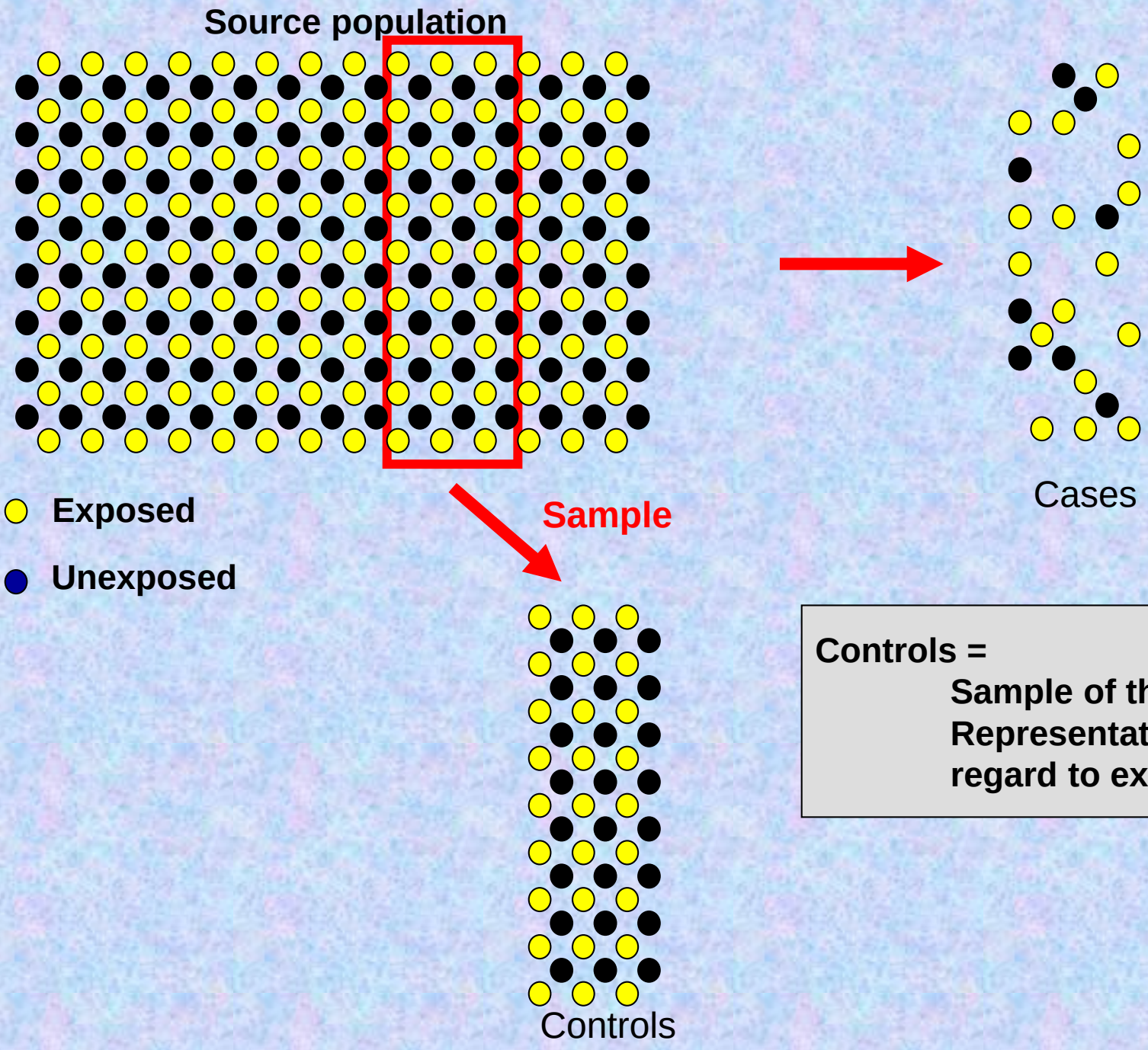


Cases

Exposed

Unexposed


Source population
Exposed
Unexposed
Sample
Cases
Controls



Controls =

Sample of the denominator
Representative with
regard to exposure

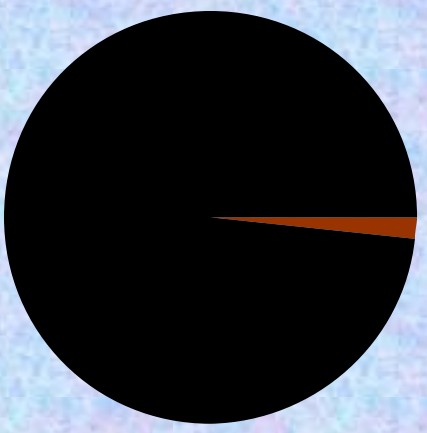
	Cases	Controls
Exposed	a	b
Not exposed	c	d
Total	a + c	b + d
% exposed	a/(a+c)	b/(b+d)

CASE-CONTROL STUDIES

- **Basic Idea:**
  - **Cases – Should represent all cases in the population**
  - **Controls – Should represent all persons without disease in the population**

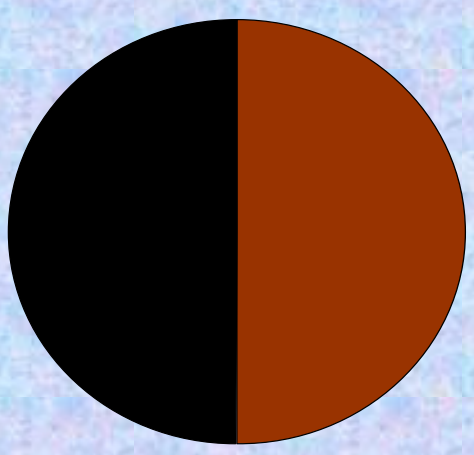
CASE-CONTROL STUDIES

Population



- Lung Cancer Cases
- Healthy

Sample



- Lung Cancer Cases
- Control

REVIEW

- A design used to assess the relationship between the exposure to a risk factor and the development of a disease
- It compares the exposure distributions between the groups of patients with and without the disease.
- It typically uses only a fraction of the subjects in the non-disease group.

Characteristics of the Design

- Retrospective
- No randomization
- Population at risk is often undefined
- Ascertainment of exposure history

Implementation a Case-Control Study: Practical Issues

- Selecting a study base representative of the intended population
- Defining the disease
- Choosing the cases and controls
- Exclusion criteria
- Ascertainment of exposure

Selection of the Study Base

- **Hospital based case-control studies: The study base is the collection of clinical records of the participating hospitals.**
  - **Berkson's Bias: Cases and controls experience different hospital admission rates.**
- **Population based case control studies: The Study base is the collection of subjects who would become cases if they develop diseases.**
  - **Neyman's Bias: Case group not representative of the intended population.**

Diagnostic Criteria and Case Selection

- **Diagnostic criteria: unambiguous definition under equal diagnostic surveillance.**
- **Sources of cases:**
  1. **Persons with the disease seen at a care facility in a specified period of time.**
  2. **Persons with the disease in a more general population in a period of time.**

Selection of Controls
Basic Principles

- **True Representation of the Study Base: The controls should be selected so that they truly represent the distribution of exposure in the study base from which the cases are selected.**
- **Comparable Accuracy: There should be no differential misclassification between the two groups.**

Selection of Controls: Sources

- The controls should be drawn from the population of which the cases represent the affected individuals.
- Sampling Frames:
  1. Population of an administrative area (eg. HMOs)
  2. Hospital patients
     1. Difference with target population
     2. Cost effective
  3. Relatives of the cases (spouses and siblings)
  4. Associates of the cases (neighbors, co-workers, etc)

Matching

- **Frequency matching**
- **Individual matching**

Matching

- **Advantages:**
  - **Sometimes the only way of control of some confounding in certain situations**
  - **Increasing power**
  - **Straightforward way to obtain a comparable group**

Matching

- **Disadvantages:**
  - Some time impossible
  - Association between matching variable and the outcome can't be assessed
  - Not possible to assess theadditive interaction between matching variable and exposure
  - Increased int validity may result in reduced ext. validity
  - Considering OVERSTIMATION: not highcorrelation between the variable of interest and matching variable
    - eg: matching ethnic background
  - No statistical power is gained if the matched variable is a weak confounder

Selection of Controls: Sampling Schemes

- Total population – no sampling
- Random and systematic sampling
- Matching – deliberately select the controls in such a way as to make them similar to the cases with respect to certain confounding variables.
- Multiple control groups.

Multiple controls

- Similar
- different

Exclusion Criteria

- Exclusion criteria should not alter the exposure rate in one of the two groups.
- Examples:
  1. Low-level lead exposure and mental retardation-children with lead related diseases were excluded from the control group;
  2. Reserpine and breast cancer-patients with thyrotoxicosis, renal disease, and cardiovascular diseases were excluded from the control group.

Information on Exposure

- The most common sources of information on exposure are patients (or parents, in the case of children).
- Other sources include relatives, hospital records, employment records, etc.
- When information is obtained via interviews, recall bias is often a concern.

Information on Exposure: Comparability and Validity

- **Comparability: If the inaccuracy in exposure reporting affects the two groups to a different degree, the study may yields questionable conclusions.**
- **Validity: The information on exposure reflects the true level of exposure.**

Advantage and disadvantages

Case control studies

- **epidemiologists use them to study a huge variety of associations.**
- **more frequently than other analytical studies**

Case control studies

Advantages:

- **Rare diseases**
- **Several exposures**
- **Long latency**
- **Rapidity**
- **Low cost**
- **Small sample size**
- **No ethical problem**
- **Efficient, cost-effective for rare outcomes**

Case control studies

Disadvantages:

- **Selection bias**
- **Measurement of exposure information**
- **Control of confounding factors**
- **Not suitable for rare exposure**
- **? Sequence of events ?**
- **Only one outcome**
- **Does not yield incidence or relative risk (although in some cases these can be inferred using external information)**
- **BIAS**

Effects

INCIDENCE or PREVALENCE
DISEASE or EXPOSURE

Intuitively

if the frequency of exposure is higher among cases than controls

then the incidence rate will probably be higher among exposed than non exposed.

Distribution of cases and controls according to exposure in a case control study

	Cases	Controls
Exposed	a	b
Not exposed	c	d
Total	a + c	b + d
% exposed	a/(a+c)	b/(b+d)

Distribution of myocardial infarction cases and controls by amount of physical activity

Physical activity	Myocardial Infarction	Controls
>= 2500 Kcal	190	230
< 2500 Kcal	176	136
Total	366	366
% exposed	51.9%	62.8 %

$$\text{Odds} = \frac{\text{Probability that an event will happen}}{\text{Probability that the even will not happen}}$$

$$\text{Odds} = \frac{\text{Probability that an event will happen}}{1 - (\text{Probability that the event will happen})}$$

Case control study

	Cases	Controls
Exposed	a	b
Not exposed	c	d
Total	a + c	b + d

Odds of exposure among cases =
Probability to be exposed among cases
Probability to be unexposed among cases

$$\text{Odds Ecases} = \frac{a/(a+c)}{c/(a+c)} = a/c$$

Odds of exposure among controls =
Probability to be exposed among controls
Probability to be unexposed among controls

$$\text{Odds Econtrols} = \frac{b/(b+d)}{d/(b+d)} = b/d$$

$$OR = \frac{a/c}{b/d} = ad/bc$$

CASE-CONTROL STUDIES

	(+)	(-)
	Case	Control
RF (+)	50	20
RF (-)	50	80

- **BASIC IDEA**
- **Is the risk factor more common among cases than controls?**

ODDS FOR CASES 50:50 = 1

ODDS FOR CONTROLS 20:80 = 0.25

ODDS RATIO = 50:50/20:80 = 1/0.25 = **4**

- **RR isn’t possible to calculate in case control study**
- **OR is calculated**
- **OR is representative of RR if:**
  - **Cases are representative**
  - **Controls are representative**
  - **Disease prevalence is rare**

CASE-CONTROL STUDIES

- **Method: Population-based**
- **Prospective case-control**
-
- **<u>Cases</u>: All incident cases of childhood (<15 yo) cancer in Denver registry, 1976-1983**
- **<u>Controls</u>: Random-digit dialing match on sex, age ± 3y**

Analytical Issues

- **Association vs Causal relationship.**

- **Adjustment of confounders:**
  1. **Matching**
  2. **Model based adjustment (regression, etc)**
  3. **Propensity score method**
  4. **A common limitation of the adjustment: cannot account for the effects of the unobserved confounders.**

Final Thoughts

- **Thoughtful design and careful implementation.**
- **Reducing biases of various kinds.**
- **The workhorse of the case-control data analysis is logistic regression.**
- **Reporting a case-control study.**

Nested case control