EpiSouth

Network for the Control of Public Health Threats
in the Mediterranean Region and South East Europe

Epidemiological analytical studies

1st Module of Outbreak investigation

Belgrade, Serbia

Marie Roseline D. Bélizaire

Based on previous EPIET presentation

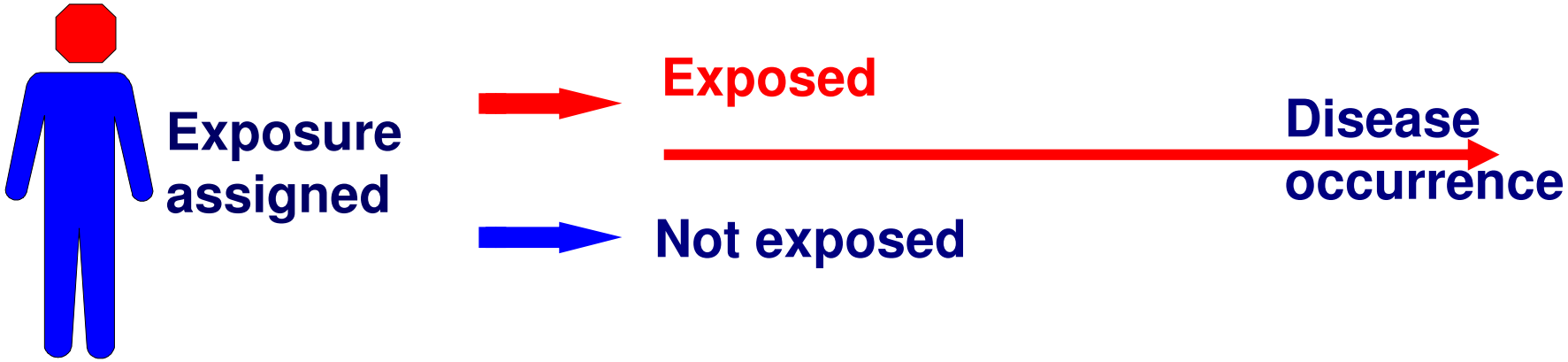
April 9, 2013

Epidemiological analytical studies

Two types

- Observation
- Experiment

Experiment



Unethical to perform experiments on people if exposure is harmful

If exposure not harmful

Treatment
Preventive measure (vaccination)

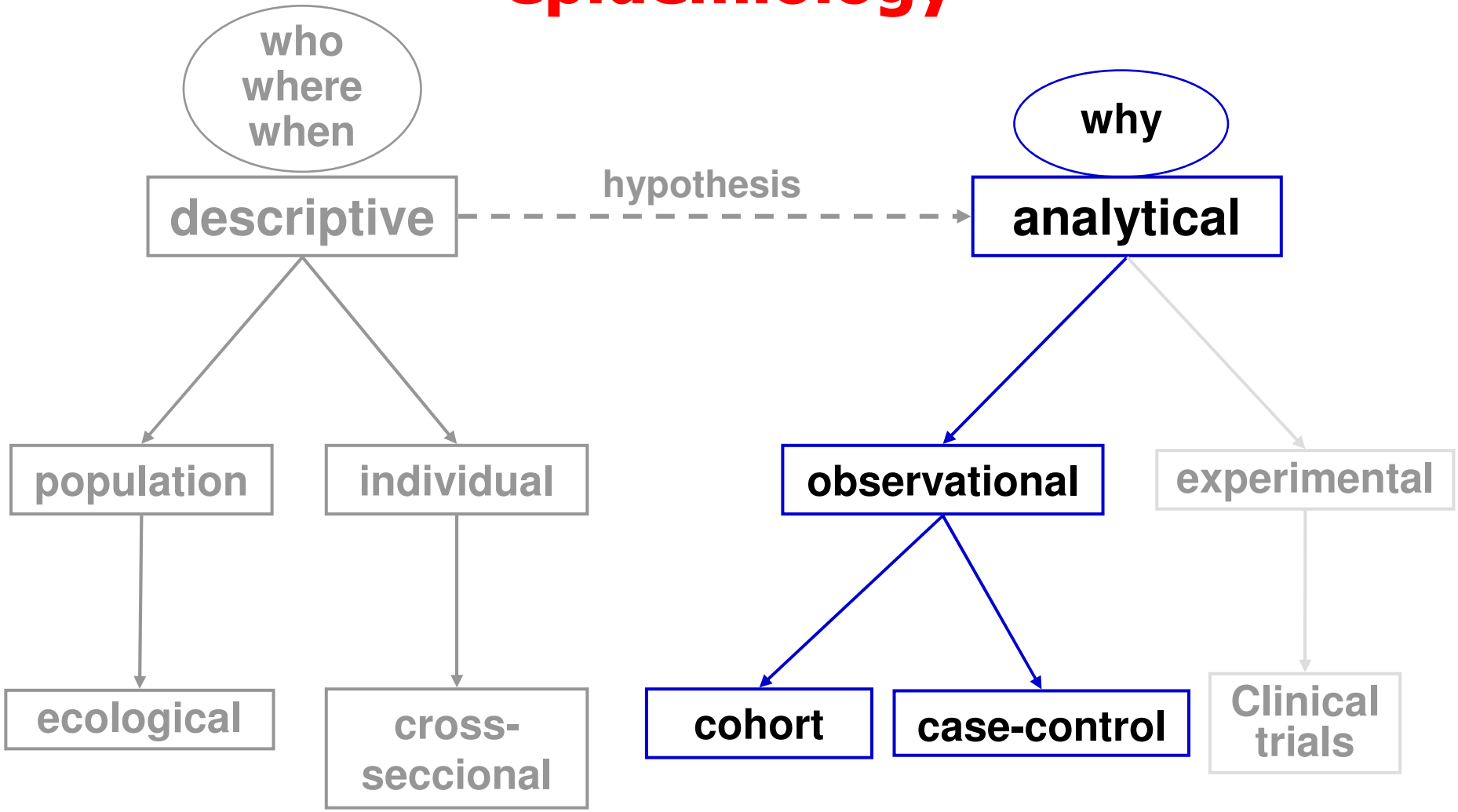
Randomised
Controlled
Trial

Blinded
Doses
Time period
Risk - effect
No bias

If RCT not possible

Left with observation of experiments designed by Nature

Descriptive studies - context in epidemiology

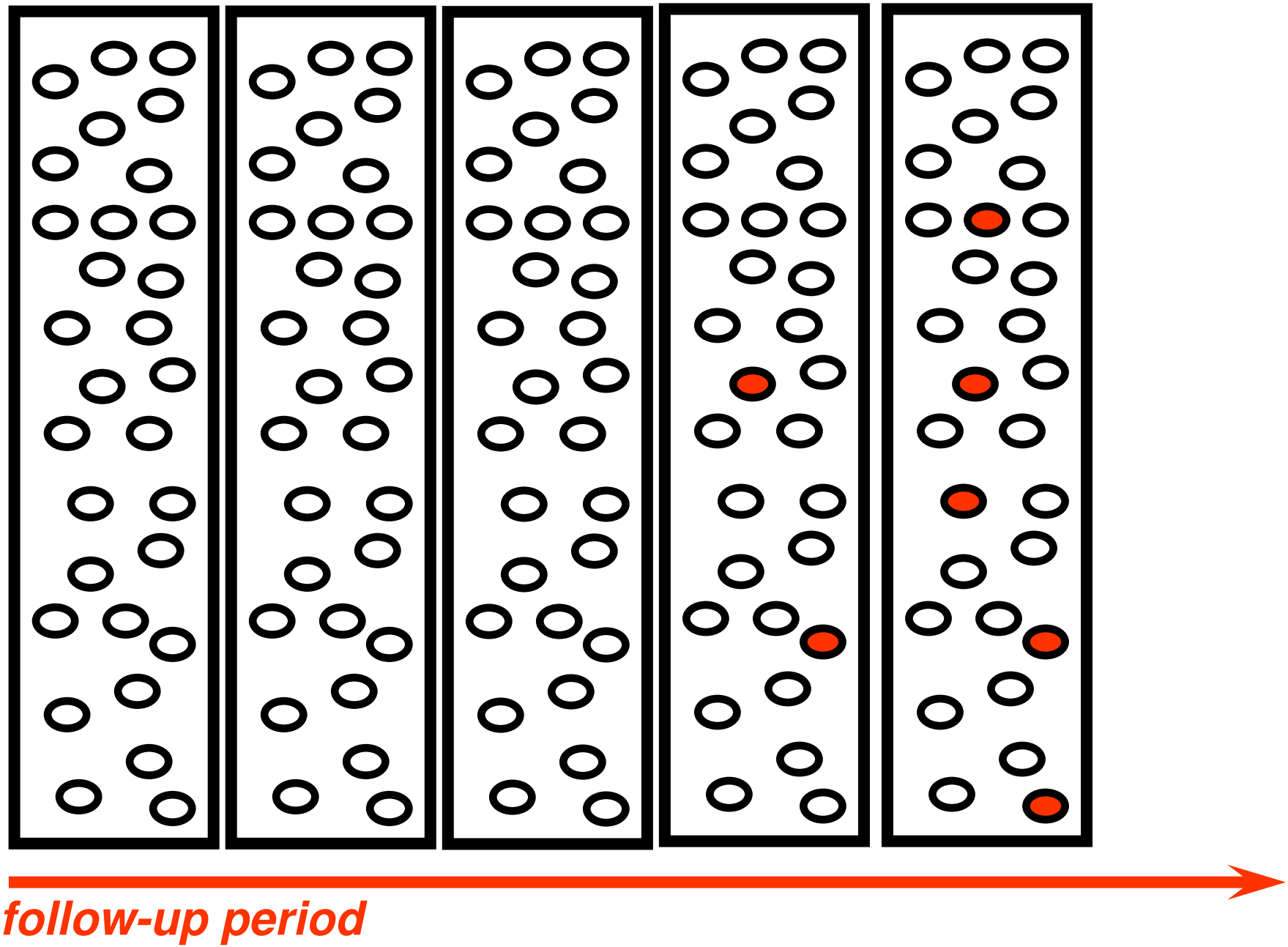


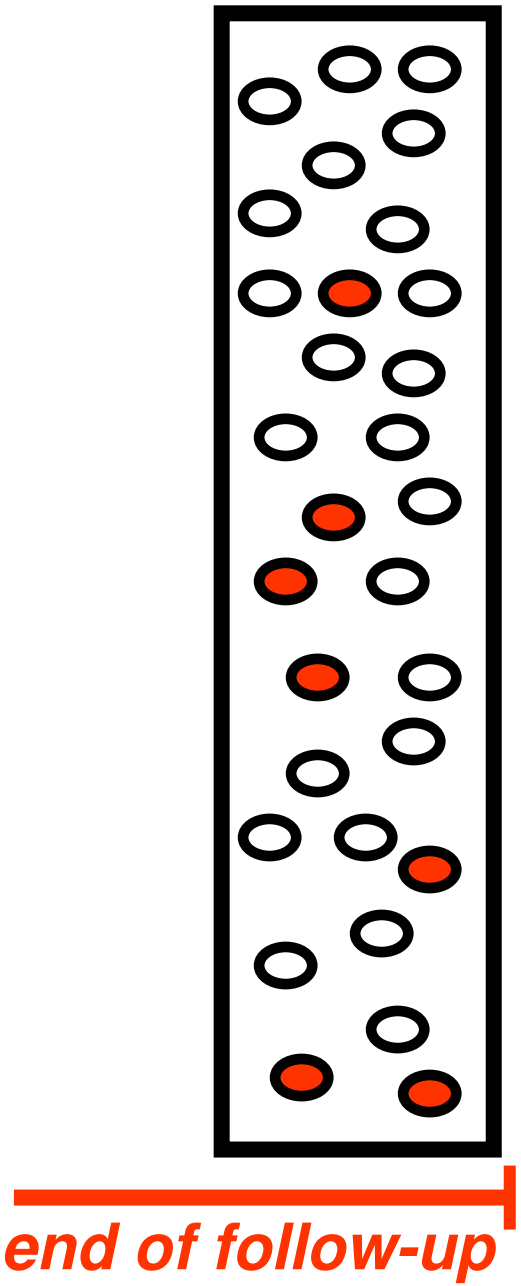
Cohort studies

marching towards outcomes

What is a cohort?

- One of 10 divisions of a Roman legion
- Group of individuals
  - sharing same experience
  - followed up for specified period of time
- Examples
  - birth cohort
  - cohort of guests at barbecue
  - occupational cohort of chemical plant workers
  - Episouth Focal point (2010-2013)


follow-up period



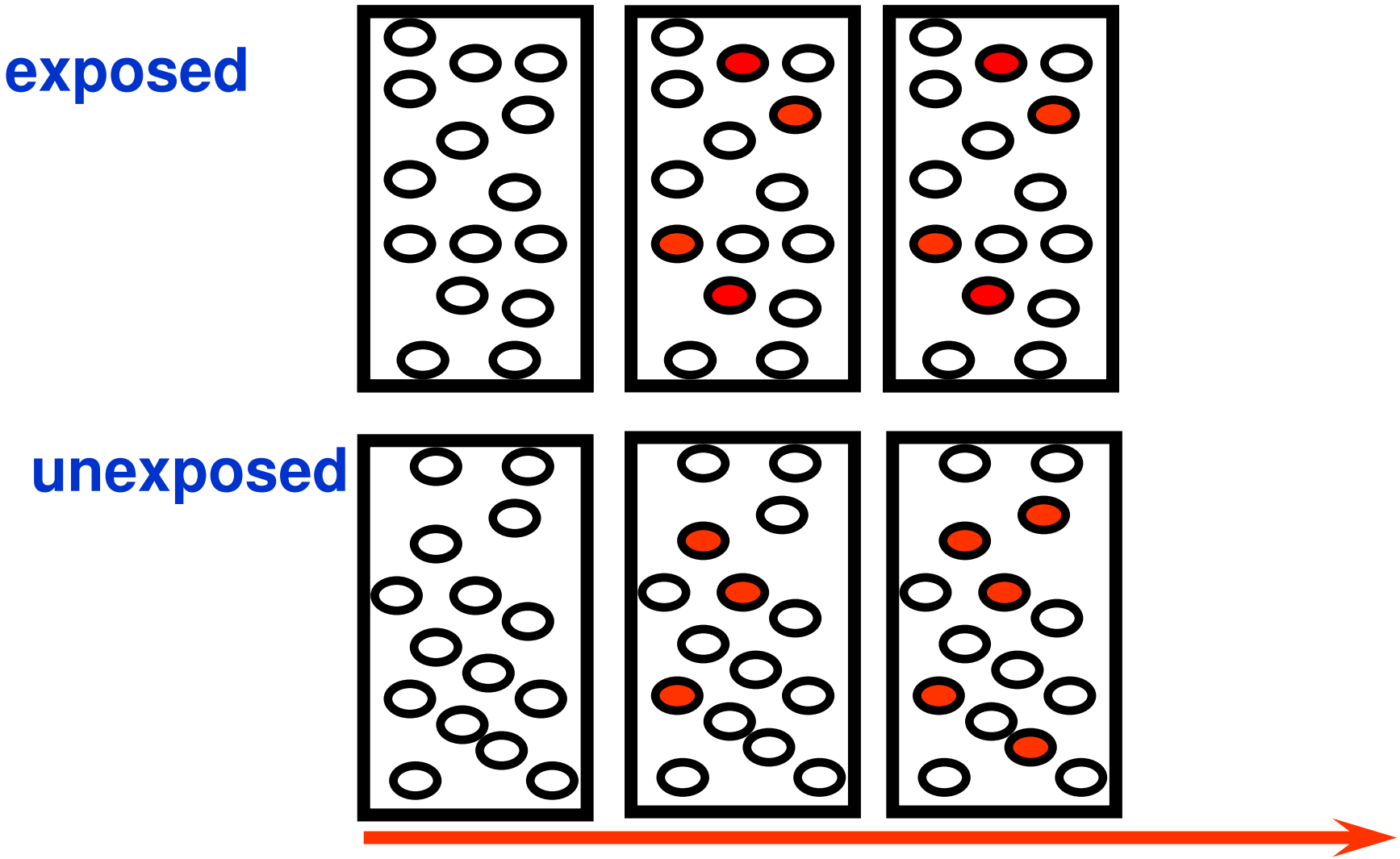
Calculate measure of frequency

- **Cumulative incidence**
  - Incidence proportion
  - Attack rate (outbreak)
- **Incidence rate**

Cohort studies

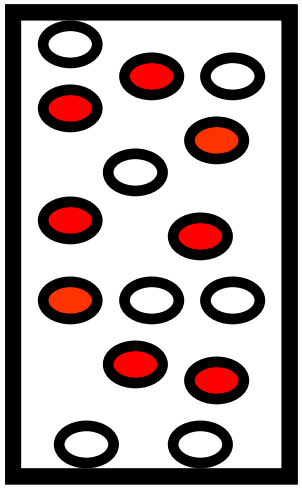
- **Purpose**
  - Study if an exposure is associated with outcome(s)?
  - Estimate risk of outcome in exposed and unexposed cohort
  - Compare risk of outcome in two cohorts
- **Cohort membership**
  - Being at risk of outcome(s) studied
  - Being alive and
  - Being free of outcome at start of follow-up

Cohort studies



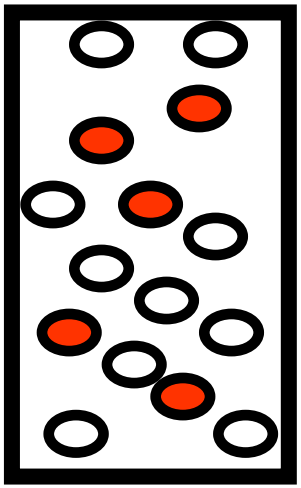
Cohort studies

exposed



Incidence among exposed

unexposed



Incidence among unexposed

Presentation of cohort data: 2x2 table

	ill	not ill	
ate ham	49	49	98
did not eat ham	4	6	10

Presentation of cohort data: Population at risk

Does HIV infection increase risk of developing TB among a population of drug users?

	Population (f/u 2 years)	Cases
HIV +	215	8
HIV -	289	1

Source: Selwyn et al., New York, 1989

Presentation of cohort data: Person-years at risk

Tobacco smoking and lung cancer, England & Wales, 1951

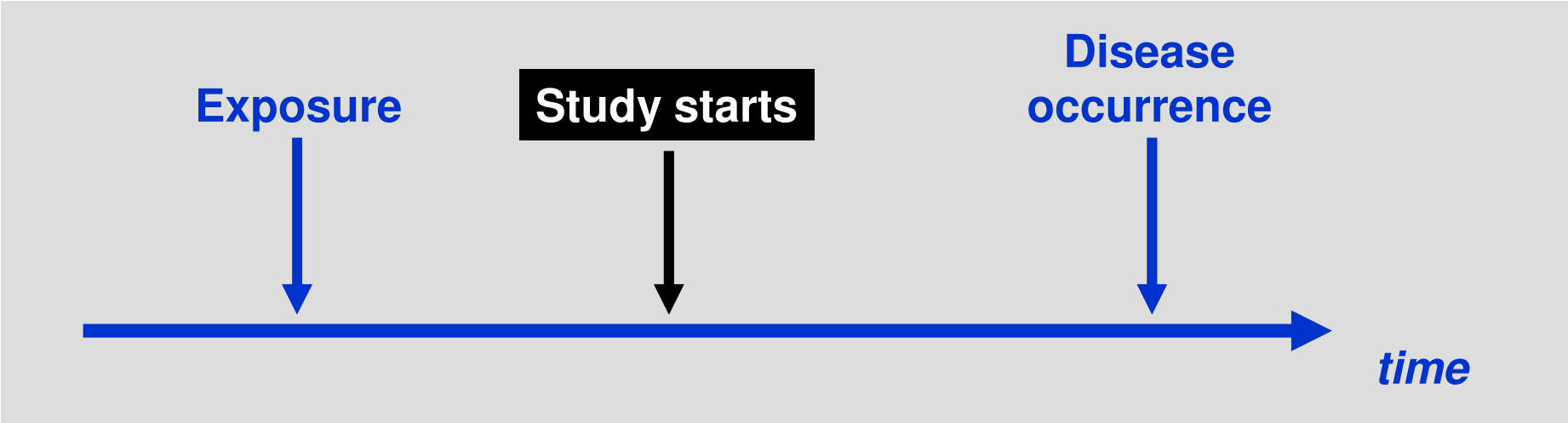
	Person-years	Cases
Smoke	102,600	133
Do not smoke	42,800	3

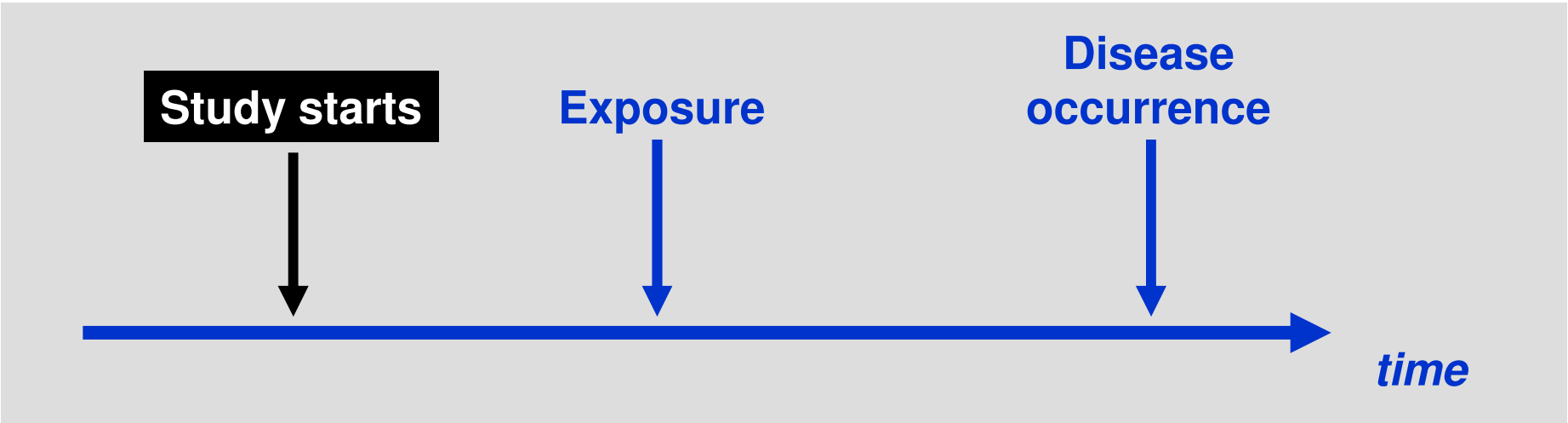
Source: Doll & Hill

Presentation of data: Various exposure levels

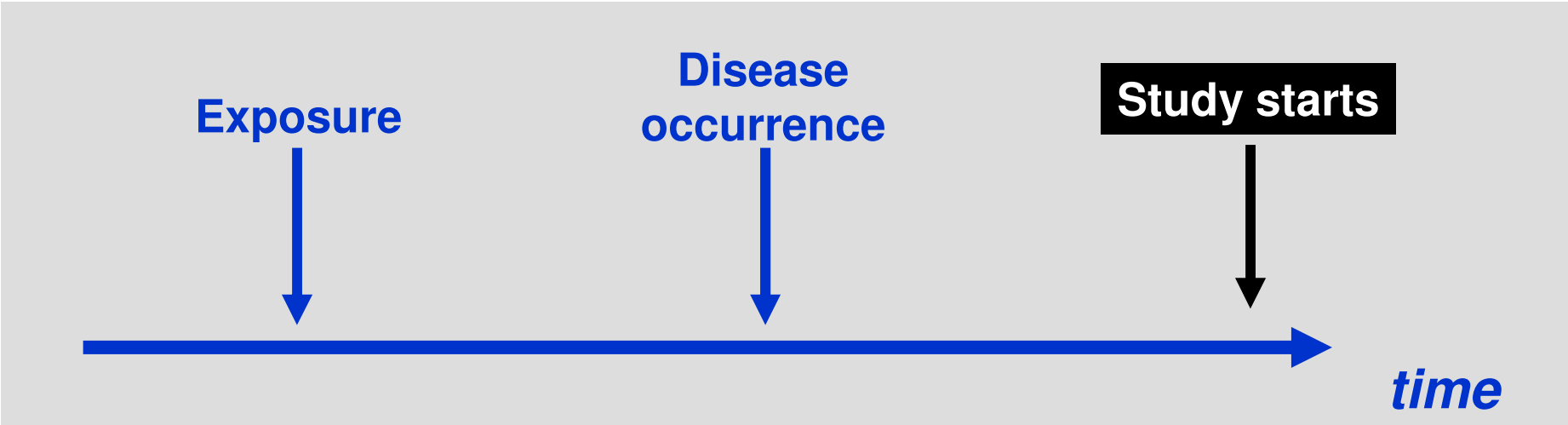
Daily number of cigarettes smoked	Person-years at risk	Lung cancer cases
> 25	25,100	57
15 - 24	38,900	54
1 - 14	38,600	22
none	42,800	3

Prospective cohort study





Retrospective cohort study



Recipe: Cohort study

- Identify group of
  - exposed subjects
  - unexposed subjects
- Follow up for disease occurrence
- Measure incidence of disease
- Compare incidence between exposed and unexposed group

Our objective is to compare

the incidence rate in the exposed population

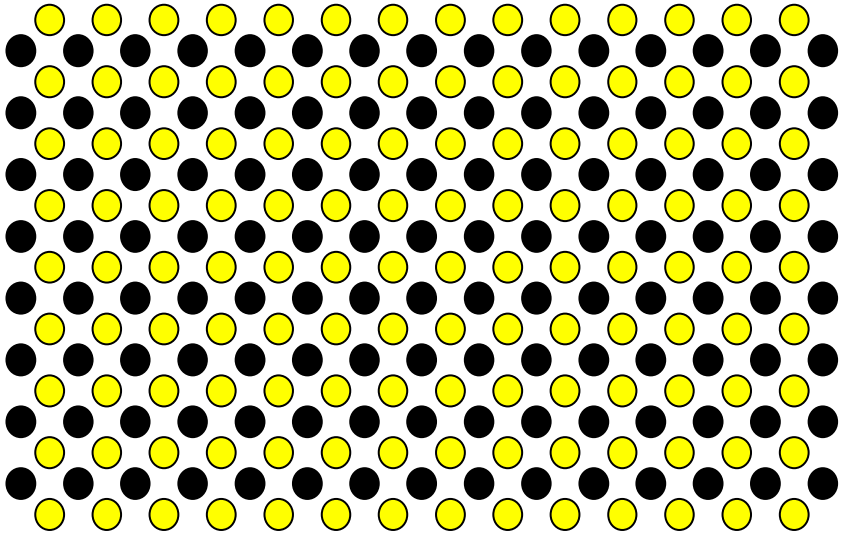
to the rate that would have been observed

in the same population, at the same time

if it had not been exposed

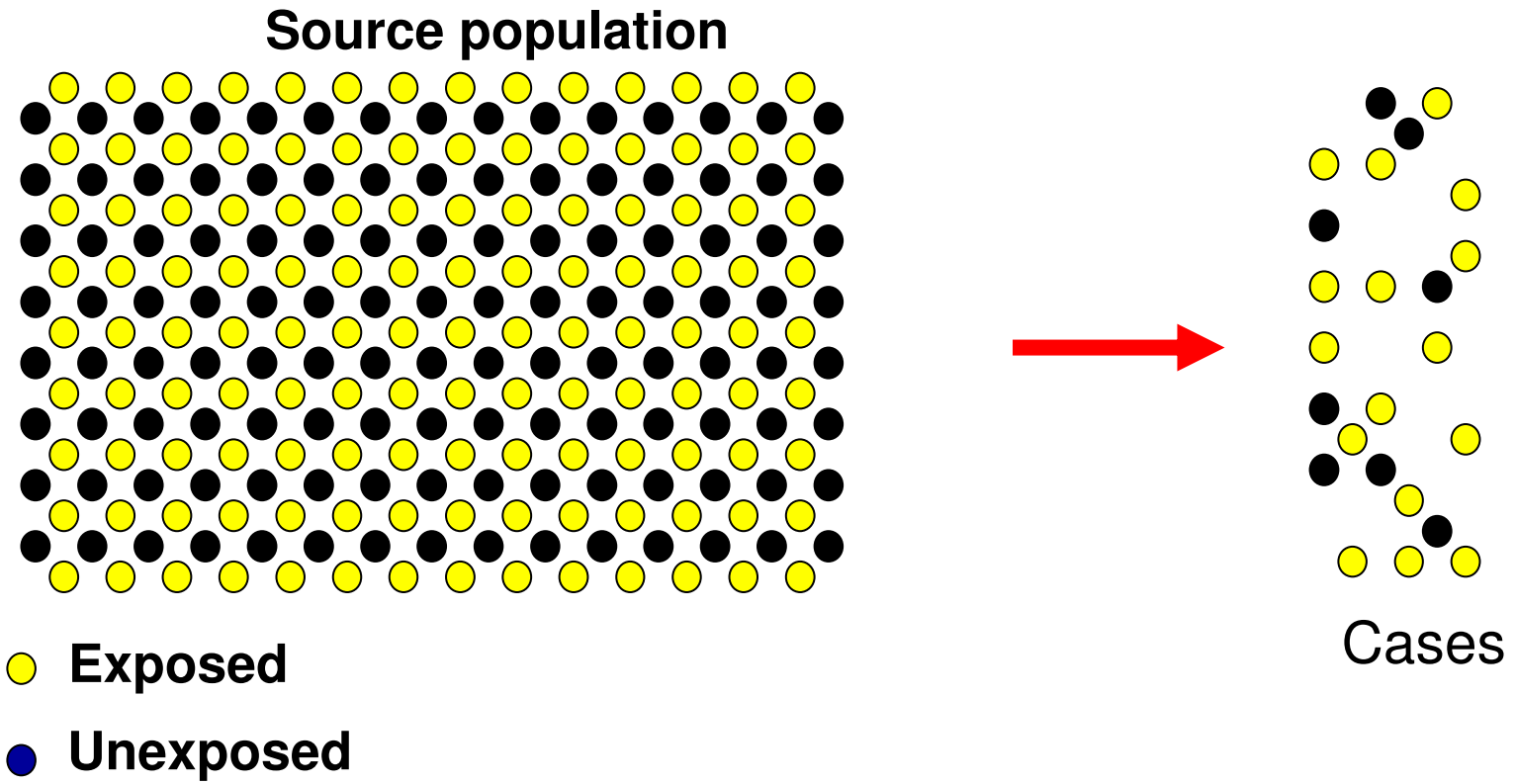
Principle of case control studies

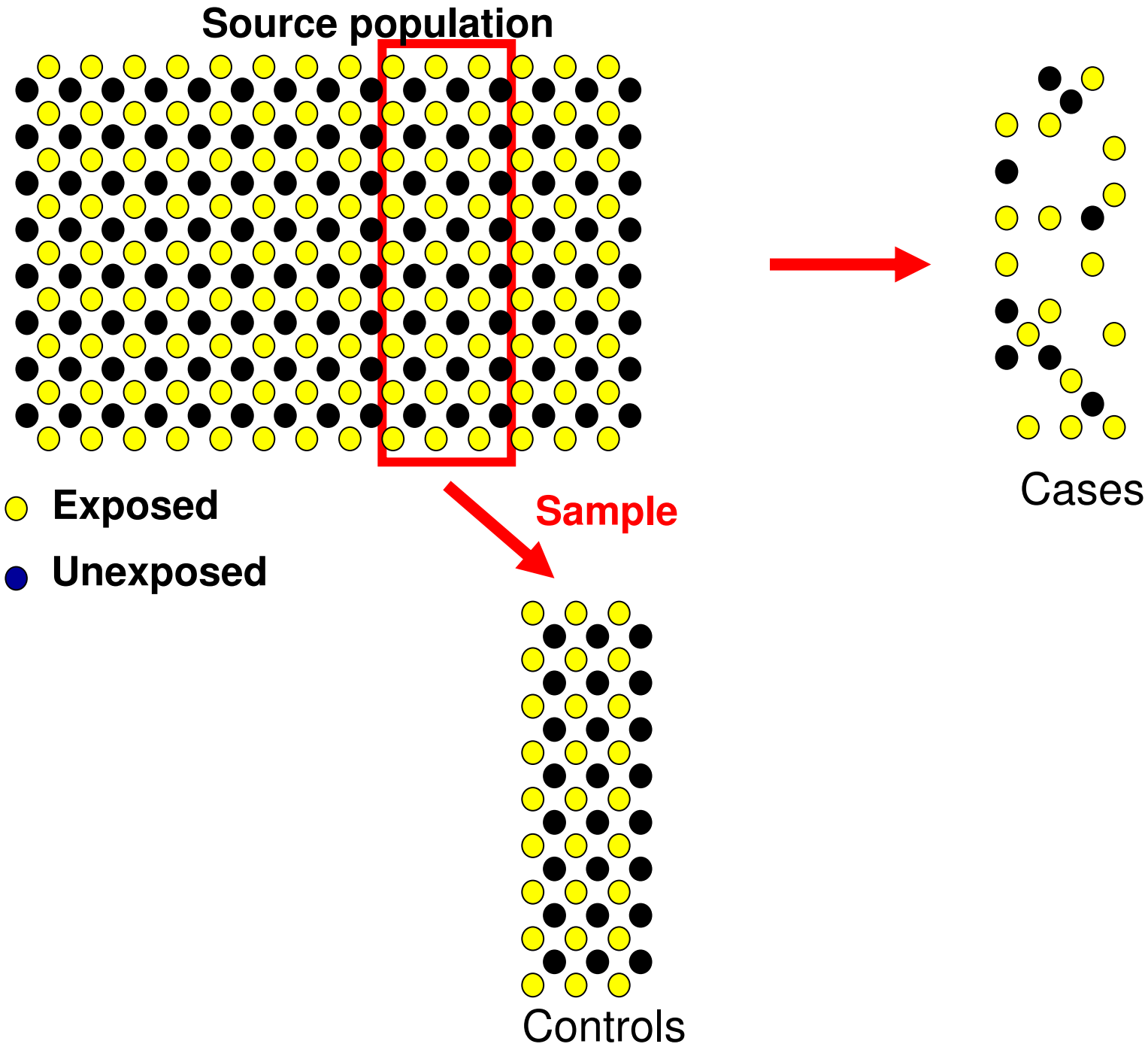
Source population

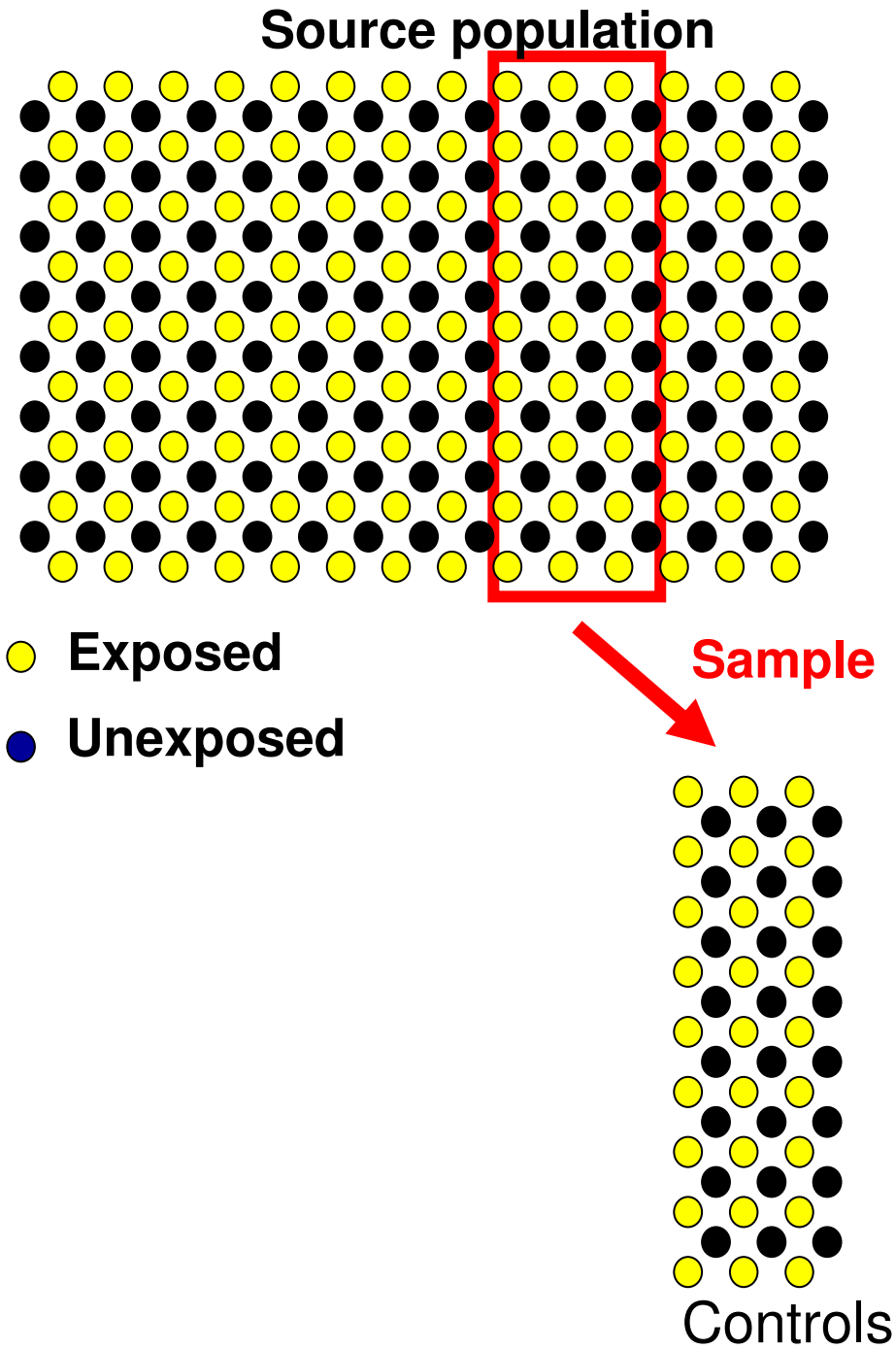


Exposed

Unexposed


Source population
Cases
Exposed
Unexposed


Source population
Exposed
Unexposed
Sample
Cases
Controls



Controls:

Sample of the denominator

Representative with regard to 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

Case control study

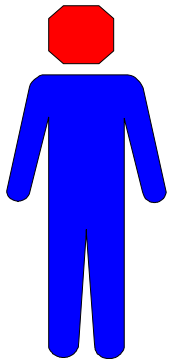
Exposure

?

?

Disease

Controls



Retrospective nature

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 by oral contraceptive use in cases and controls

Oral contraceptives	Myocardial Infarction	Controls
Yes	693	320
No	307	680
Total	1000	1000
% exposed	69.3%	32%

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

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

Volvo factory, Sweden, 3000 employees,
Cohort study
200 cases of gastroenteritis

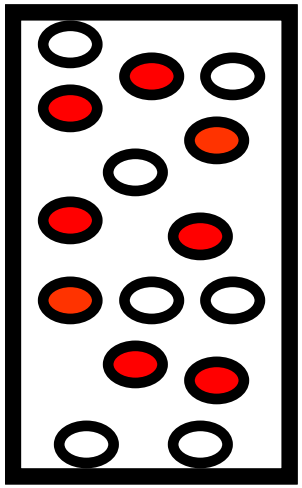
Water Consumption	Cases	Controls
YES	150	?
NO	50	?
Total	200	200

Two types of case control studies

- **Exploratory**
  - New disease
  - New risk factors
  - Several exposures
  - "Fishing expedition"
- **Analytical**
  - Define a single hypothesis
  - Dose response

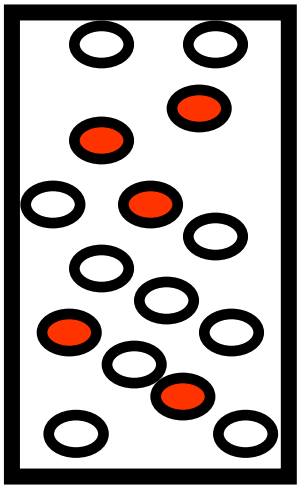
Cohort studies

exposed



Incidence among exposed

unexposed



Incidence among unexposed

Effect measures in cohort studies

- **Absolute measures**
  - Risk difference (RD) $I_e - I_{ue}$

- **Relative measures** $\frac{I_e}{I_{ue}}$
  - Relative risk (RR)
    - Rate ratio
    - Risk ratio

I_e = incidence in exposed
I_{ue} = incidence in unexposed

	ill	not ill		Incidence
ate ham	49	49	98	50 %
did not eat ham	4	6	10	40 %

Risk difference 50% - 40% = 10%

Relative risk 50% / 40% = 1.25

Does HIV infection increase risk of developing TB among drug users?

Exposure	Population (f/u 2 years)	Cases	Incidence (%)	Relative Risk
HIV +	215	8	3.7	11
HIV -	298	1	0.3	

Vaccine efficacy (VE)

Status	Pop.	Cases	Cases per 1,000	RR
Vaccinated	301,545	150	0.49	0.28
Unvaccinated	298,655	515	1.72	Ref.
Total	600,200	665	1.11	

$$VE = 1 - RR = 1 - 0.28$$

$$= 72\%$$

Various exposure levels

Exposure level	Population at risk	Cases	Incidence
High	N_1	a_1	I_1
Medium	N_2	a_2	I_2
Low	N_3	a_3	I_3
Unexposed	N_{ne}	c	I_{ue}

Various exposure levels

Exposure level	Population at risk	Cases	Incidence	RR
High	N_1	a_1	I_1	RR_1
Medium	N_2	a_2	I_2	RR_2
Low	N_3	a_3	I_3	RR_3
Unexposed	N_{ne}	c	I_{ue}	Reference

Cohort study:
Tobacco smoking and lung cancer, England & Wales, 1951

Cigarettes smoked/d	Person-years at risk	Cases	Rate per 1000 p-y	Rate ratio
> 25	25,100	57	2.27	32.4
15 - 24	38,900	54	1.39	19.8
1 - 14	38,600	22	0.57	8.1
none	42,800	3	0.07	Ref.

Source: Doll & Hill

A cohort study allows to calculate indicators which have a clear, precise meaning.

The results are immediately understandable.

Cohort studies

- Rate
- Rate difference
- Rate Ratio (strength of association)

Case control studies

- No calculation of rates
- Proportion of exposure

Any way of estimating Rate ratio ?

	Cases	Population denominator
E	a	P_1
$\overline{E}$	c	P_0

$$I_1 = a / P_1$$

$$I_0 = c / P_0$$

$$I_1 / I_0 = \frac{a/P_1}{c/P_0}$$

	Cases	Population sample
E	a	$P_1/10$
$\overline{E}$	c	$P_0/10$

$$I_1 = \frac{a}{P_1/10}$$

$$I_0 = \frac{c}{P_0/10}$$

$$I_1 / I_0 = \frac{a/P_1}{c/P_0}$$

Source population

	Cases	Pop.
E	a	P_1
$\overline{E}$	c	P_0

$I_1 = a / P_1$

$I_0 = c / P_0$

$$I_1 / I_0 = \frac{a/P_1}{c/P_0}$$

= sample

	Cases	Controls
E	a	b
$\overline{E}$	c	d

$$\frac{P_1}{P_0} = \frac{b}{d}$$

Source population

	Cases	Pop.
E	a	P_1
$\overline{E}$	c	P_0

$I_1 = a / P_1$

$I_0 = c / P_0$

= sample

	Cases	Controls
E	a	b
$\overline{E}$	c	d

$$\frac{P_1}{P_0} = \frac{b}{d}$$

$$I_1 / I_0 = \frac{a/P_1}{c/P_0} = \frac{a \cdot P_0}{c \cdot P_1} = \frac{a \cdot d}{c \cdot b} = \frac{a / c}{b / d}$$

Since $d/b = P_0 / P_1$

Case control study design

	Cases	Controls
E	a	b
$\overline{E}$	c	d

Odds ratio

$$\frac{a}{c} \Big/ \frac{b}{d} = \frac{a \times d}{b \times c}$$

Distribution of myocardial infarction by recent oral contraceptive use in cases and controls

Oral contraceptives	Myocardial Infarction	Controls	OR
Yes	693	320	4.8
No	307	680	Ref.
Total	1000	1000	
% exposed	69.3%	32%	

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

Physical activity	Myocardial Infarction	Controls	OR
>= 2500 Kca l	190	230	0.64
< 2500 Kcal	176	136	Ref.
Total	366	366	
% exposed	51.9%	62.8%	

Distribution of cases of endometrial cancer by oestrogen use in cases and controls

Oestrogen use	Cases	Controls	Odds ratio
High	a1	b1	a1d/b1c
Low	a2	b2	a2d/b2c
None	c	d	Reference

Relation of hepatocellular adenoma to duration of oral contraceptive use in 79 cases and 220 controls

Months of OC use	Cases	Controls	Odds ratio
0-12	7	121	Ref.
13-36	11	49	3.9
37-60	20	23	15.0
61-84	21	20	18.1
>= 85	20	7	49.7
Total	79	220	

Source: Rooks et al. 1979

Limitations of case-control studies

- Cannot compute directly relative risk
- Not suitable for rare exposure
- Temporal relationship exposure-disease difficult to establish
- Biases +++
  - control selection
  - recall biases when collecting data
- Loss of precision due to sampling

Disadvantages of cohort studies

- Large sample size
- Latency period
- Cost
- Time-consuming
- Loss to follow-up
- Exposure can change
- Multiple exposure = difficult
- Ethical considerations

Advantages of case control studies

- Rare diseases
- Several exposures
- Long latency
- Rapidity
- Low cost
- Small sample size
- Available data
- No ethical problem

Strengths of cohort studies

- Can directly measure
  - incidence in exposed and unexposed groups
  - true relative risk
- Well suited for rare exposure
- Temporal relationship exposure-disease is clear
- Less subject to selection biases
  - outcome not known (prospective)

Strengths of cohort studies

- Can examine multiple effects for a single exposure

Population		Outcome 1	Outcome 2	Outcome 3
exposed	N_e	I_{e1}	I_{e2}	I_{e3}
unexposed	N_{ne}	I_{ue1}	I_{ue2}	I_{ue3}
		RR_1	RR_2	RR_3

The cohort study is the gold standard of analytical epidemiology

CASE-CONTROL STUDIES HAVE THEIR PLACE IN EPIDEMIOLOGY **but if cohort study possible, do not settle for second best**