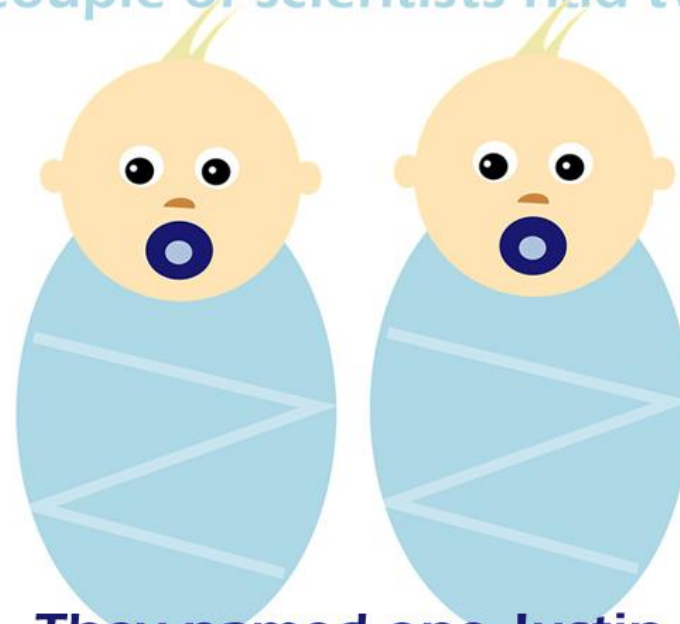


A couple of scientists had twins



They named one Justin  
and the other Control

Author: Conf. Dr. Cosmina-Ioana Bondor

# Lecture 3 – Study protocol, Case-control study



ALWAYS



SEEK



KNOWLEDGE

# Case-control study

# Scenario

- Results: Of the 100 children **with** nonsyndromic **oral clefts**, 40 declared **maternal alcohol intake** above standard limits
- Of the 100 patients **without** nonsyndromic oral clefts, 10 declared maternal alcohol intake above standard limits

# Case – control studies

Study a **link** between  
factor  
disease occurrence

= Evaluation of the frequency of a **factor** in a group of patients with the **disease** compared to a group of controls free of the disease

Selection

- presence/absence of the disease

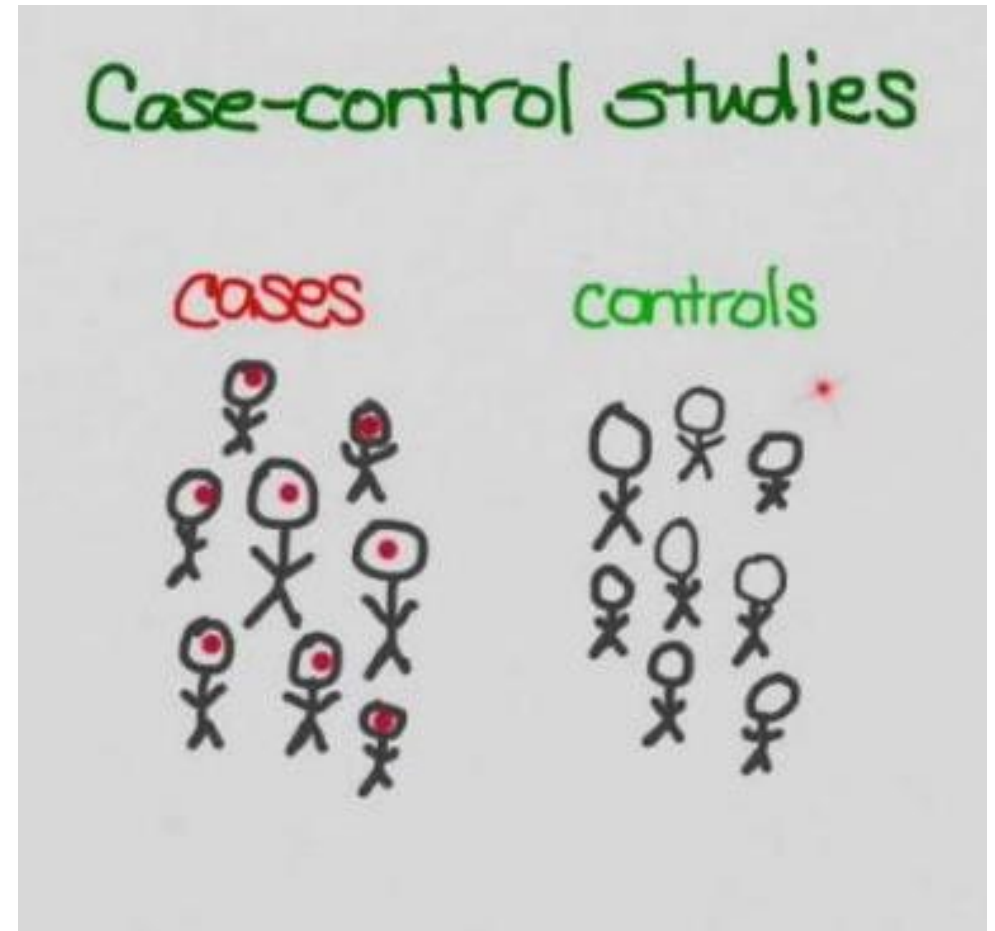
2 groups:

cases - with disease X

controls - without disease X

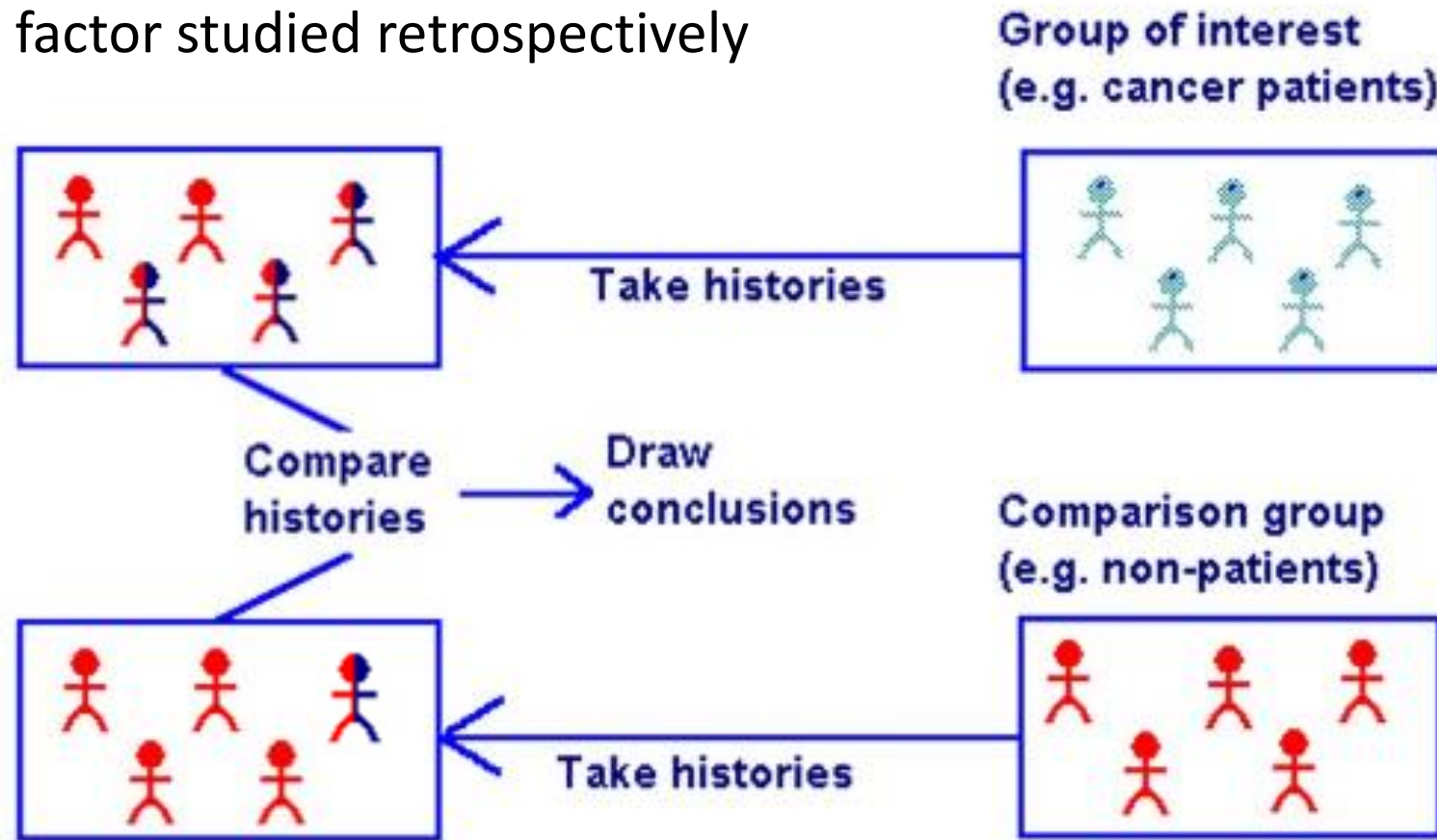
selected from the same population as the cases,  
randomly, so that we do not have selection errors

The factor studied retrospectively



# Case Control Studies

The factor studied retrospectively



Present = here we start the study

# Contingency table – Case – control studies

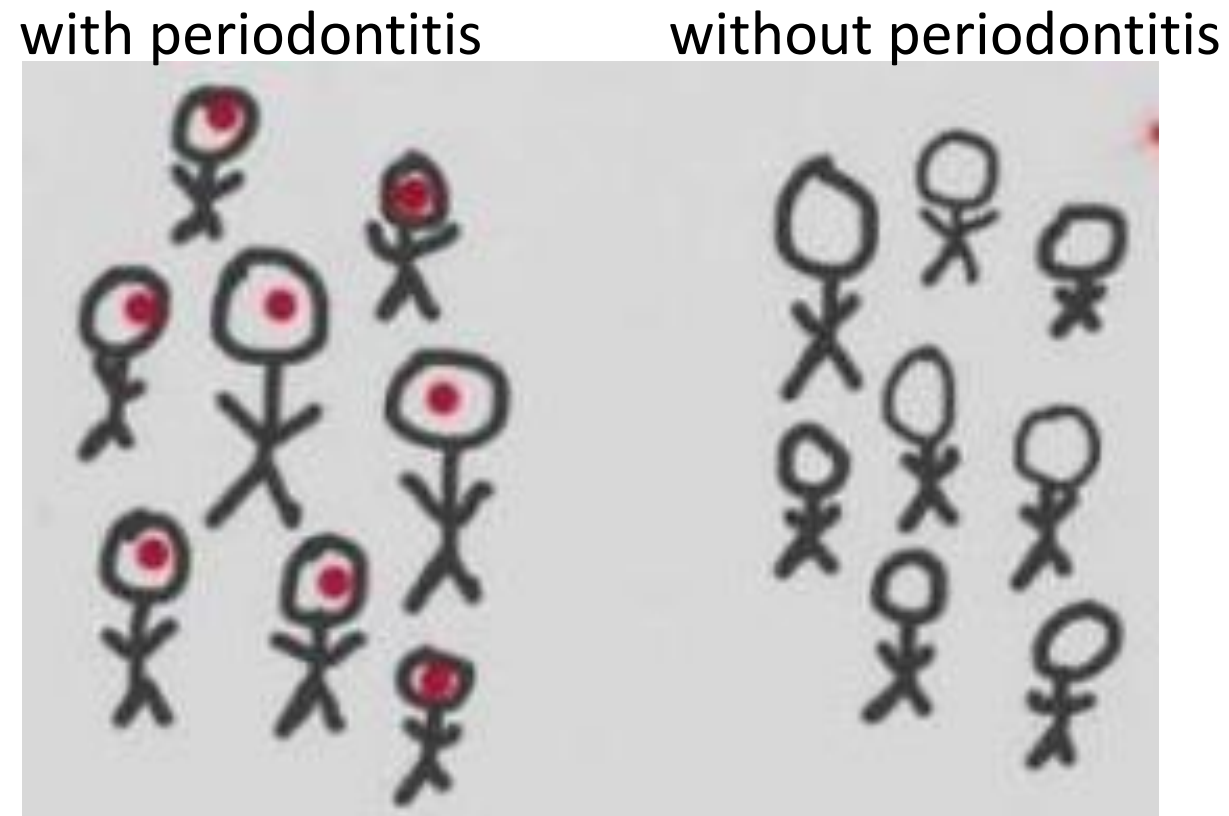
	Disease <sup>+</sup>	Disease <sup>-</sup>
Factor <sup>+</sup>	a	b
Factor <sup>-</sup>	c	d
	Total Disease <sup>+</sup>	Total Disease <sup>-</sup>

Disease<sup>+</sup> - Disease present, Disease<sup>-</sup> - Disease absent, Factor<sup>+</sup> - factor present, Factor<sup>-</sup> - factor absent

! Calculations only in the component on the right of the line or on the left of the line

# Case-control study: Is there an association between smoking and periodontitis?

- 2 groups of patients
  - with periodontitis
  - without periodontitis
- obesity was noted retrospective
  - from the patients file



# Case – control studies

## Objectives

demonstrate the existence of a link between the disease under study and a possible prognostic factor

## possible results

- the risk factor was present in a statistically significant **higher** percentage of patients with the disease compared to those without the disease
- the risk factor was present in a statistically significant **smaller** percentage of patients with the disease compared to those without the disease
- the risk factor was **not** present in a statistically significant **different** percentage of patients with the disease compared to those without the disease

quantify the link

how strong is the association?



# Risk factors

**Exposed**

400

**Non-exposed**

600

**Exposed**

100

**Non-exposed**

900

1000

**Case**

1000

**Controls**

**Present – recruiting  
patients into the study**

**Time**

**Direction of investigation**

More percentages of ill people in case group

	Case	Control
	Periodontitis <sup>+</sup>	Periodontitis <sup>-</sup>
Obesity <sup>+</sup>	400	100
Obesity <sup>-</sup>	600	900
Total	1000	1000

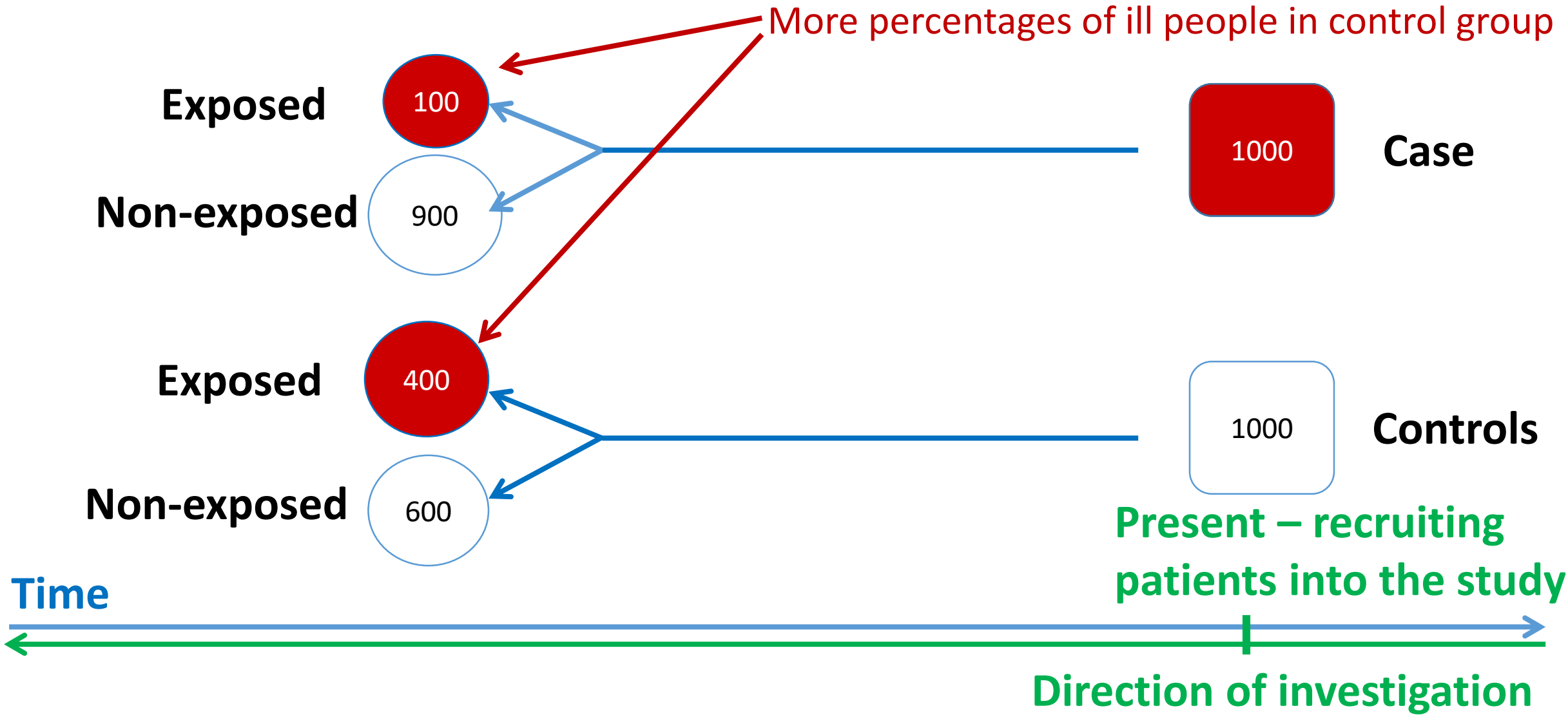
2000 subjects aged 60 years.

Group 1: with Periodontitis

Group 2: without Periodontitis

data on obesity are collected retrospectively from the record

# Protective factor



# Case-control study advantages

- Low costs
- Relatively short duration
- Several factors can be studied at the same time
- Useful
  - in the study of **rare** pathologies
  - in the case of a **long time between exposure and the onset** of the disease

# Disadvantages of Case-control studies

- Cannot determine cause of disease
- Cannot calculate:
  - prevalence of disease in population
  - relative risk of disease in case of exposure
- can study only one objective (outcome, disease)
- risk of sampling errors
  - loss of deaths from the study
- risk of observational errors
  - biased recall
  - forgetting

Data analysis

# Contingency table – Case – control studies

	Disease <sup>+</sup>	Disease <sup>-</sup>
Factor <sup>+</sup>	a	b
Factor <sup>-</sup>	c	d
	Total Disease <sup>+</sup>	Total Disease <sup>-</sup>

Disease<sup>+</sup> - Disease present, Disease<sup>-</sup> - Disease absent, Factor<sup>+</sup> - factor present, Factor<sup>-</sup> - factor absent

! Calculations only in the component on the right of the line or on the left of the line

	Case	Control
	Periodontitis <sup>+</sup>	Periodontitis <sup>-</sup>
Obesity <sup>+</sup>	400	100
Obesity <sup>-</sup>	600	900
Total	1000	1000

2000 subjects aged 60 years.

Group 1: with Periodontitis

Group 2: without Periodontitis

data on obesity are collected retrospectively from the record



# Statistical tests for the contingency table

- To show the existence of an association between a factor and a disease
- The **Chi-square test** tests the association between two qualitative variables
- If <20% of the cells in the theoretical (expected) table have values <5, the **Fisher exact test** is used
  - $p < 0.05$  means that there is a significant association between the risk factor and the disease
  - $p \geq 0.05$  means that there is NO significant association between the risk factor and the disease
- ! it does not result from the study which is the cause and which is the effect

Study objective: obesity and periodontitis are dependent

	Case	Controls
	Periodontitis <sup>+</sup>	Periodontitis <sup>-</sup>
Obese <sup>+</sup>	400	100
Obese <sup>-</sup>	600	900
Total	1000	1000

40% obese

10% obese

Null hypothesis: we assume by absurdity that obesity and osteoporosis are independent  
we apply the Chi-square Test

test result:  $p=0.0000001$  ( $p<0.05$  - we have statistically significant differences)

Conclusion: we reject the null hypothesis, we accept the alternative hypothesis (the objective of the study) obesity and osteoporosis are dependent

# Statistic for quantifying the factor-disease link

- Odds ratio:

$$OR = \frac{a*d}{b*c}$$

B – disease, F - factor

	B <sup>+</sup>	B <sup>-</sup>	
F <sup>+</sup>	a	b	a+b
F <sup>-</sup>	c	d	c+d
	Total B <sup>+</sup>	Total B <sup>-</sup>	Total=n

# Clinical interpretation - OR

- $OR = 1$  – there is no relationship
- $OR \approx 1$  – unimportant relationship (small)
- $OR > 3$  – risk – important relationship (big)
- $OR \approx 0$  – protection – important relationship (big)
- Interpretation - statistical point of view:
  - $OR = 1.3$ 
    - 1.3 times higher chance of illness for those who are exposed compared to those who are not
    - minor importance

# Generalization of OR to the entire population

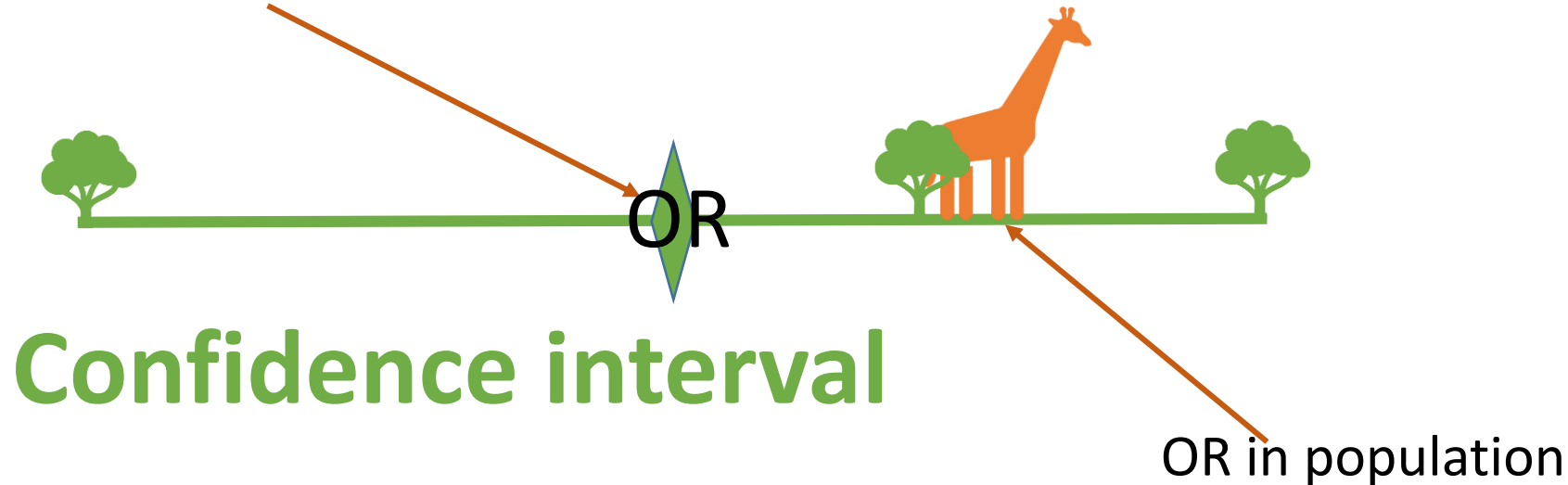
- calculating 95% **confidence interval**
  - using online applications
- Ex. OR=1.3; 95% CI (1.1 – 4.2)

OR calculated on the sample



95% confidence interval for OR

**Estimating a fix value** = OR calculated from the sample



**Confidence interval**

- estimated with an interval
- population OR is in the estimated interval with 95% probability
- can be anywhere in the estimated interval, we cannot tell where is in the interval,
  - if we want to be more precise we will have to repeat the study in a bigger sample → estimate a narrow 95% CI
- there is 5% level of error = probability of 5% that the population OR is not in the interval

# A possible interpretation of the confidence interval

OR="value"; 95% CI ("lower limit"; "upper limit")

## 1. Interpretation

- in the population OR
  - can be say with a 95% probability
    - it is between the "lower limit" and the "upper limit"
  - What it mean:
    - don't know exactly what the OR is in the population
    - would know if we did the study on the entire population,

## 2. Interpretation

- in the population OR
  - can be say with a 5% error
    - it is somewhere in the confidence interval

# A possible interpretation of the confidence interval

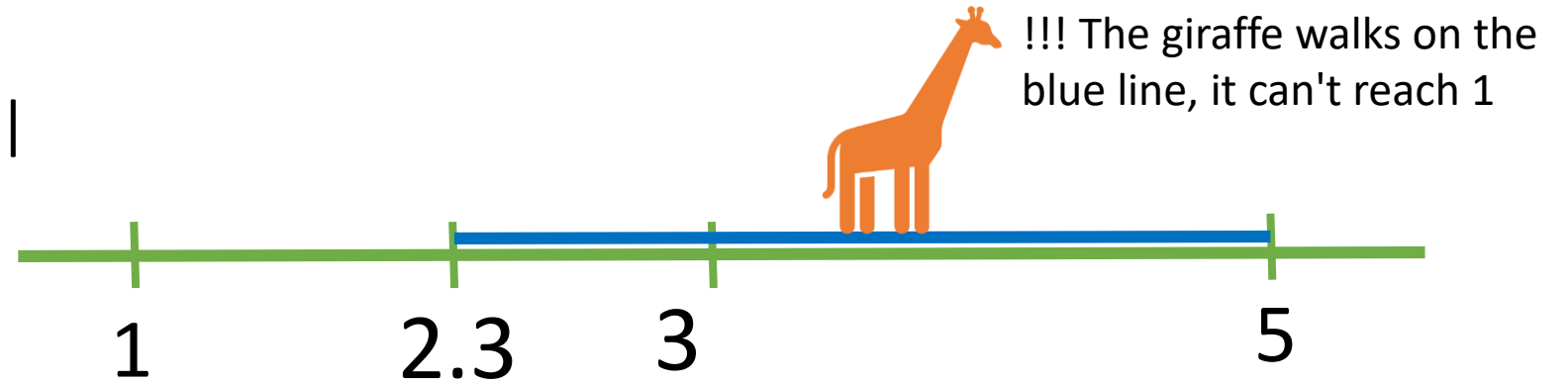
We have also the associated p-value?

## 3. Interpretation

- If  $p$  is statistically insignificant ( $p \geq 0.05$ ), then  $OR=1$  is in the CI
  - possible in the population  $OR=1 \rightarrow$  the factor and the disease are not associated
  - e.g.  $p=0.08$ ;  $OR=1.3$ ; 95% CI (0.9; 1.8)  $1 \in (0.9; 1.8)$
- If  $p$  is statistically significant ( $p < 0.05$ ) then  $OR=1$  is not in the CI
  - in the population  $OR \neq 1 \rightarrow$  the factor and the disease are associated
  - e.g.  $p=0.03$ ;  $OR=1.3$ ; 95% CI (1.1; 1.8)  $1 \notin (1.1; 1.8)$
- !!!  $OR = 1$  – there is no relationship



## Confidence interval



### 4. Interpretation (if we do not know p-value)

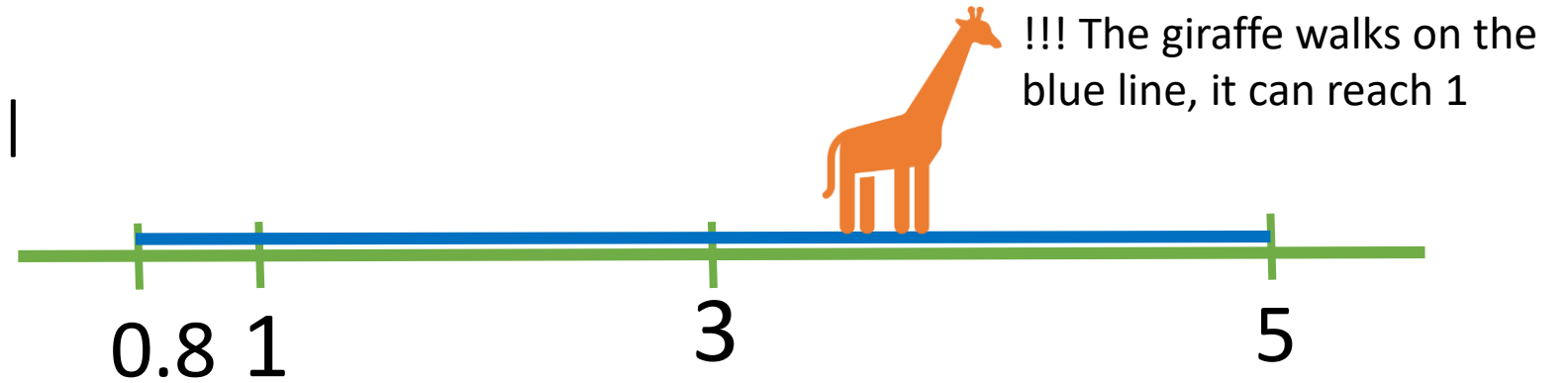
- OR = 3; 95% IC (2.3; 5)

$$1 \notin (2.3; 5) \leftrightarrow p < 0.05$$

→ there is a statistically significant relationship between the factor and the disease;  
The chance of disease is 3 times higher in those exposed than in those not exposed

if OR=1 there is no relationship

## Confidence interval



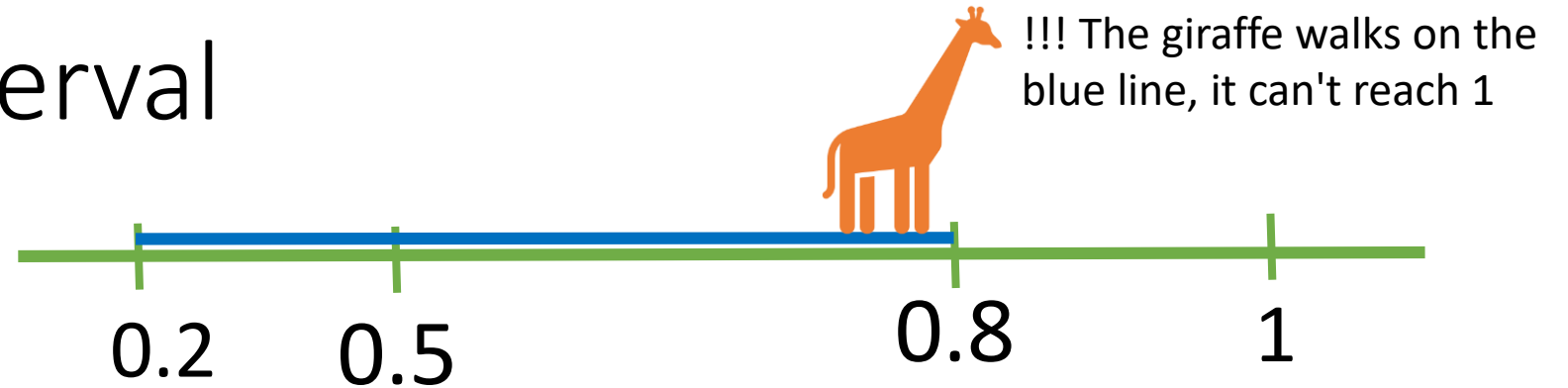
### 4. Interpretation (if we do not know p-value)

- OR=3; 95% IC (0.8; 5)

$$1 \in (0.8; 5) \leftrightarrow p \geq 0.05$$

→ there is no statistically significant relationship between the factor and the disease;  
We cannot say that there is a risk of disease in those exposed to the factor compared to those not exposed

# Confidence interval



- 4. Interpretation (if we do not know p-value)
- OR=0.5; 95% IC (0.2; 0.8)
  - $1 \notin (0.2; 0.8) \rightarrow p < 0.05$
  - there is a statistically significant relationship between the factor and the disease;
  - The risk of disease is double in those **not exposed** compared to those exposed (**protective factor**)

# A possible interpretation of the confidence interval:

## 5. Clinical interpretation

- wide interval
  - e.g. (1.1 – 27)
  - imprecise study
- narrow interval
  - e.g. (2.2 – 2.6)
  - precise study
- you have to appreciate
  - the range is wide, so the study is imprecise
  - the range is narrow, so the study is precise

# A possible interpretation of the confidence interval:

## 6. Clinical interpretation

- lower margin much greater than 1
  - e.g. (4.2 – 11)
  - important risk factor
- lower margin close to 1 and upper margin close to 1
  - e.g. (1.2 – 1.8)
  - unimportant risk factor
- lower margin close to 1 and upper margin far from 1
  - e.g. (1.2 – 11)
  - risk factor of unclear importance
- you appreciate
  - just argue well

# A possible interpretation of the confidence interval:

## 6. Clinical interpretation

- upper margin much less than 1
    - e.g. (0.2 – 0.25)
    - significant protective factor
  - upper margin close to 1 and lower margin close to 1
    - e.g. (0.8 – 0.95)
    - unimportant risk factor
  - upper margin close to 1 and lower margin far from 1
    - e.g. (0.2 – 0.95)
    - risk factor of unclear importance
- you appreciate
    - just argue well

# The problem of assessing causality (is the risk factor the cause of the disease?)

- A single observational (case-control) study is not enough to establish that the factor studied is the cause of the disease!
- We need:
  - more studies
  - experimental studies

# Scenario

- Results: Of the 100 patients with gangrene of the foot, 40 declared alcohol consumption above standard limits
- Of the 100 patients without gangrene, 10 declared alcohol consumption above standard limits

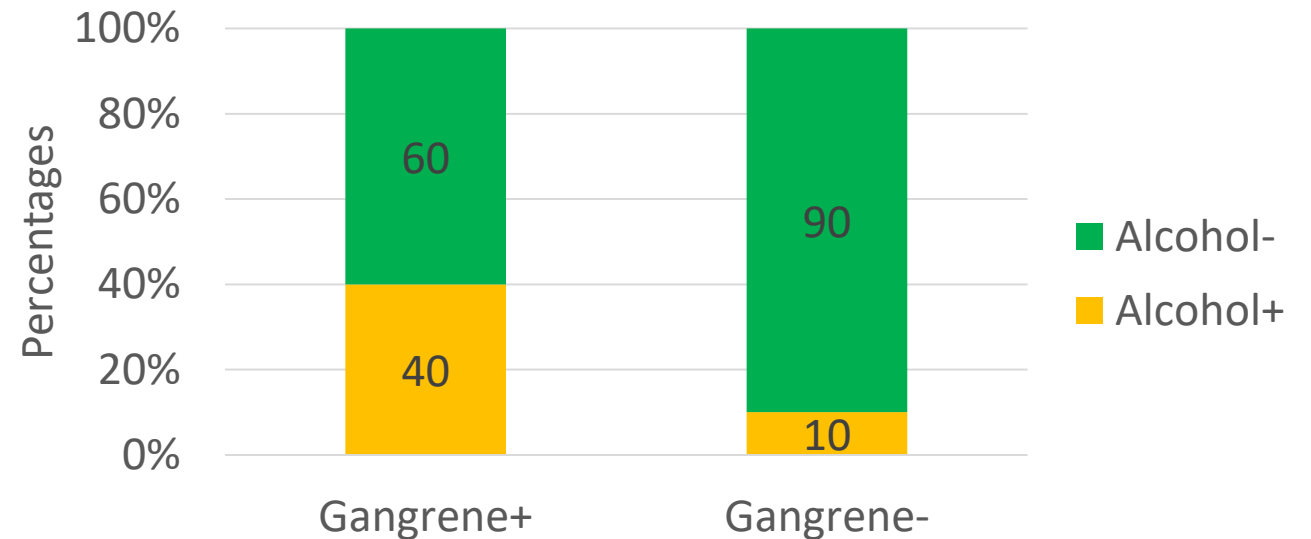


# Observed contingency table

	Gangrene <sup>+</sup>	Gangrene <sup>-</sup>
Alcohol <sup>+</sup>	40	10
Alcohol <sup>-</sup>	60	90
Total	100	100

Gangrene<sup>+</sup> - patients with gangrene, Gangrene<sup>-</sup> - patients without gangrene,  
Alcohol<sup>+</sup> - patients who consume alcohol above the standard limit, Alcohol<sup>-</sup> -  
patients who do not consume alcohol above the standard limit

- Frequency of alcohol consumption in those with gangrene=0.40  
versus
- Frequency of alcohol consumption in those without gangrene=0.10
- 40% of patients with gangrene consume alcohol
- 10% of patients without gangrene consume alcohol
  - Compare 40% with 10%



# Chi-square test

- Null hypothesis ( $H_0$ ): Alcohol consumption and foot gangrene are independent
- Alternative hypothesis ( $H_1$ ): Alcohol consumption and foot gangrene are dependent
  - $p=0.000001$
  - $p<0.05$  reject  $H_0$ , accept  $H_1$ : Alcohol consumption and foot gangrene are dependent

# OR – Odds ratio

$$\bullet \text{ OR} = \frac{\frac{\frac{a}{a+c}}{\frac{b}{b+d}}}{\frac{\frac{c}{a+c}}{\frac{d}{b+d}}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc} = \frac{40 \cdot 90}{60 \cdot 10} = 6$$

- The odds is 6 times higher for a subject with gangrene to consume alcohol above the standard limit than for a subject without gangrene
- The odds ratio indicates an important risk factor (has a high value)

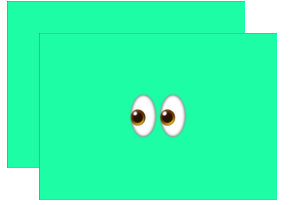
# 95% Confidence interval for OR

- OR=6, 95% CI (2.5 – 10)

The 95% confidence interval indicates the presence of a risk factor (the value 1 is not in the interval)

- The interval is wide – the study is imprecise
- Both ends of the CI indicate clinically important value – the association is relatively clinically important

# Question



- <https://app.wooclap.com/BFKAN2C4?from=event-page>

# Example

- Smoking and lung cancer
- After the case-control study conducted in 1950 and after numerous other case-control studies conducted



- Controversies: another factor present in smokers may cause cancer



- 1956 Doll conducts a study on a cohort of doctors (41,024)
- Mortality rate due to lung cancer in “heavy-smokers” compared to “smokers” 20 times higher
- The rate increased with the amount of cigarettes smoked/day
- Mortality of smokers compared to those who quit smoking: 3 times higher

DOLL R, HILL AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. Br Med J. 1956 Nov 10;2(5001):1071-81.

Thank you!