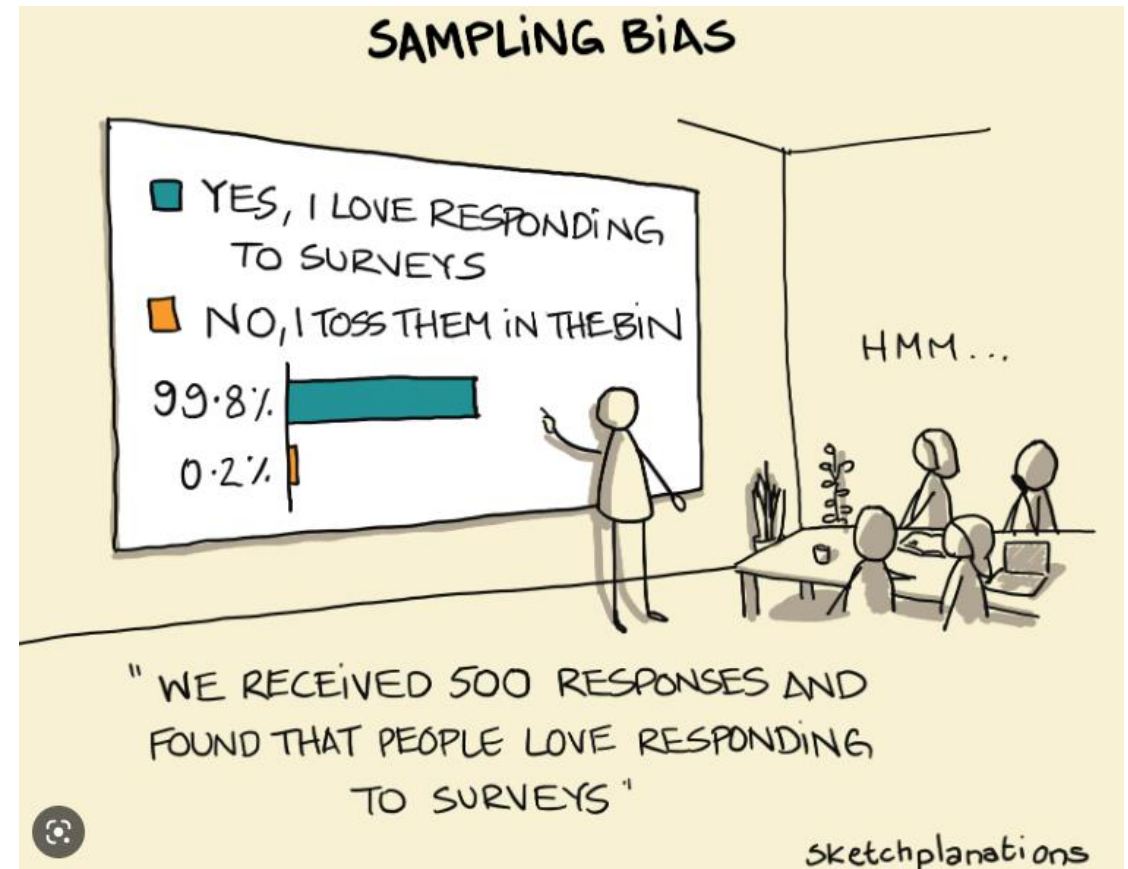


Author: Conf. Dr. Cosmina-Ioana Bondor

Lecture 4 – Cohort study



- A** ALWAYS
- S** SEEK
- K** KNOWLEDGE

Objective

- By the end of this lecture, students should be able to:
 - Define a **cohort study**
 - Distinguish **prospective vs retrospective cohorts**
 - Understand **incidence** and **risk estimation**
 - Calculate **relative risk**
 - Identify **bias and confounding**
 - Interpret cohort studies in **dental research**

Key question:

- **Does exposure lead to disease?**
- Example:
- **Does obesity increase the risk of periodontitis?**

Study type

- Observational
 - Cross-sectional
 - Case-control
 - **Cohort**
- Experimental
 - Randomized clinical trial

This lecture → Cohort studies

Cohort

- Cohort - definition
 - a group of subjects
- Purpose
 - studying whether exposure to a factor influences the occurrence of disease
 - the cohort is followed over a period of time

Exposure → Follow-up → Outcome

- Ex:
 - Sugar consumption → development of dental caries

Cohort studies

- Random sample from the population (cohort).
- share a common characteristic or exposure

- Ex.
- Children born in the same year
- Smokers vs non-smokers
- Patients with dental implants

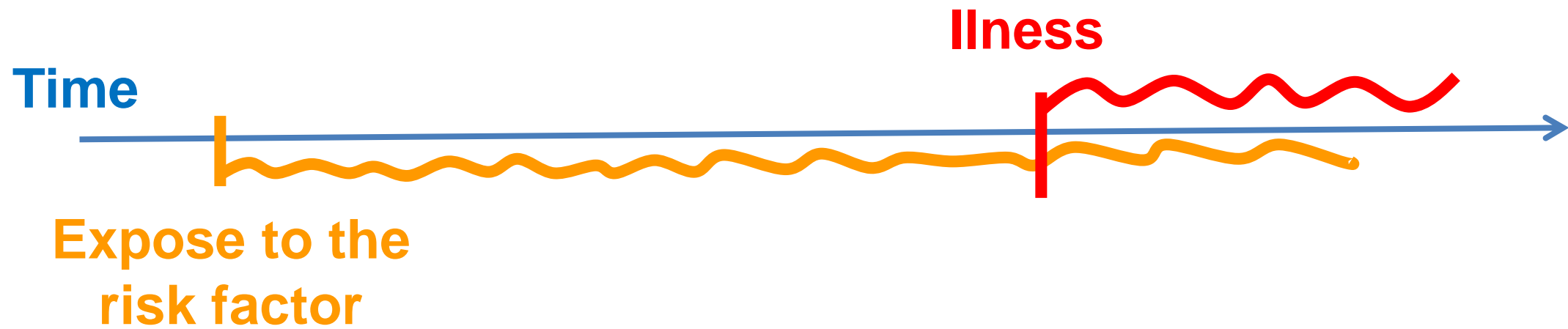
Cohort studies

- Start with individuals **without disease**
- Divided into:
 - Exposed group – the studied factor present
 - Non-exposed group – the studied factor absent
- Factor
 - Risk
 - Protective
 - PrognosisFactor

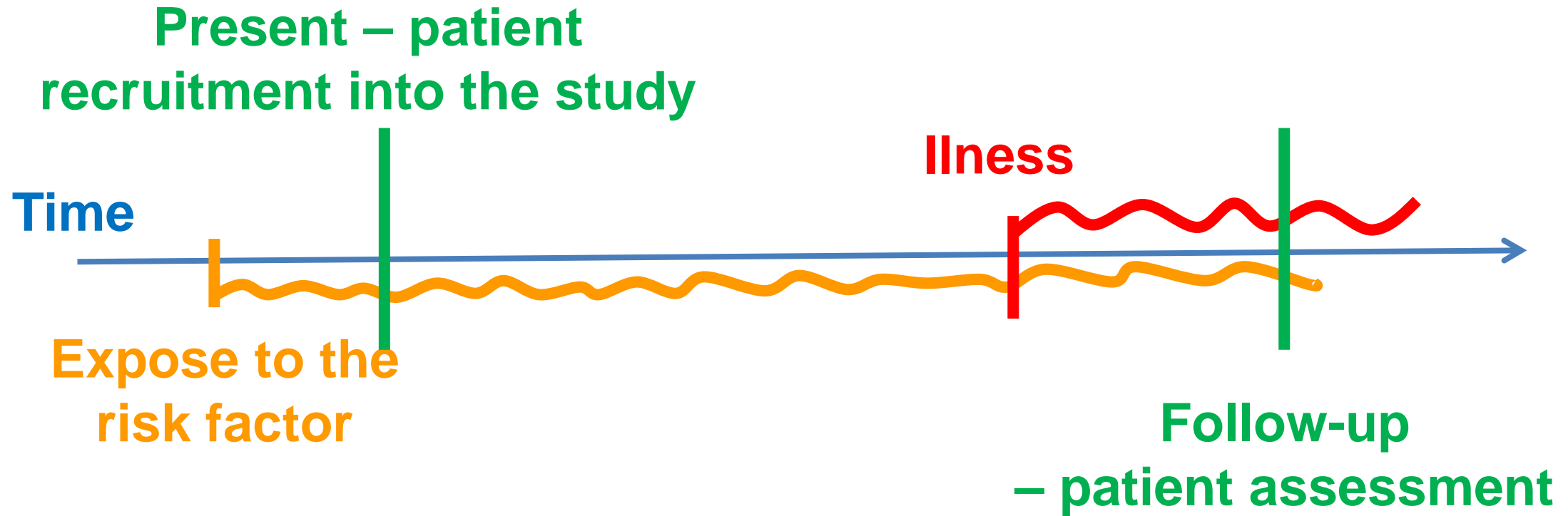


Follow over time to observe:
Disease occurrence

Cohort study

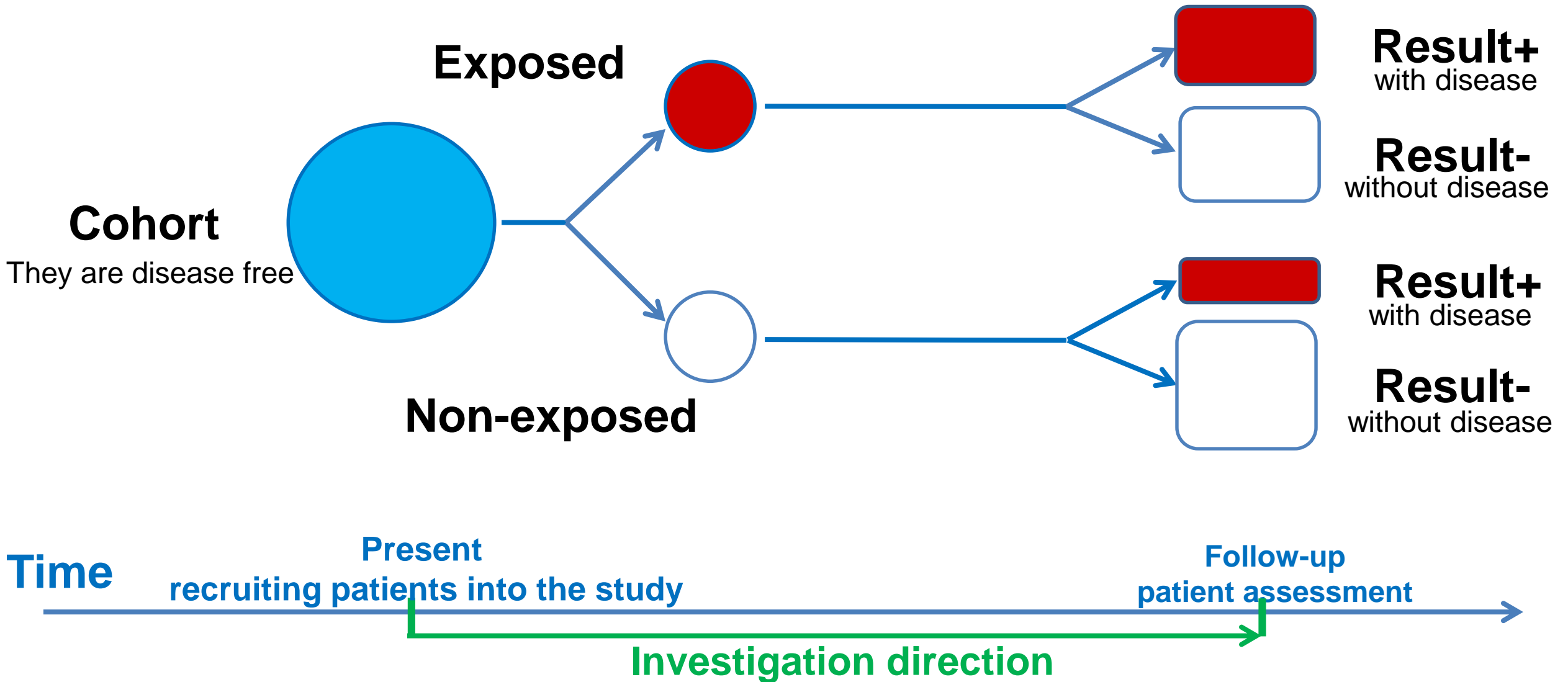


Cohort study



Cohort study

- The researcher recruits subjects
 - exposed disease-free
 - non-exposed disease-free
- !!! excluded from the study:
 - those with the disease being studied
- presence } study factor → tracking time → onset of the disease
- absence }



Contingency table – Expose-unexposed

	Disease ⁺	Disease ⁻	
Factor ⁺	a	b	a+b
Factor ⁻	c	d	c+d

Disease⁺ - Present disease, Disease⁻ - Absent disease, Factor⁺ - present factor, Factor⁻ - absent factor

Prospective cohort study

- random sample from the studied population
- Exposure measured **before disease occurs**
- Then participants are **followed into the future**

- Example:
 - Include dental students → Evaluate smoking
 - Follow for 10 years → Evaluate periodontal disease
- Advantages:
 - better data quality
 - clear temporal relationship

Retrospective cohort study

- Uses **existing data**
- Exposure and outcomes already occurred.
- Example sources:
 - hospital records
 - dental registries
 - insurance databases
- Advantages:
 - faster
 - cheaper
- Limitations:
 - incomplete data

Cohort studies

- Japanese university students
- **body mass index (BMI)** is related to **bleeding on probing?**
 - one group of students **with increase BMI** periodontitis free
 - one group of students **without increase BMI** periodontitis free
- Evaluate **bleeding on probing**
 - when they **start the studies**
 - when they **finish the studies**

Cohort studies

- Participants
 - ≥ 40 years of age
- periodontitis influenced rheumatoid arthritis
 - group with periodontitis rheumatoid arthritis free
 - group without periodontitis rheumatoid arthritis free
- groups were matched by sex, age, and household income at a 1:1 ratio.
- the 2 groups ($n = 691,506$) were followed one year
- rheumatoid arthritis evaluated
 - at the inclusion (start)
 - after one year (finish)

Measures of Disease Frequency

- Important concept:
- Incidence
- Incidence = number of new cases in a population over time
- Example:
- 20 new cases of periodontitis among 200 patients over 5 years.
- Incidence = 10%

	Disease ⁺	Disease ⁻	
Factor ⁺	a	b	Total Factor ⁺
Factor ⁻	c	d	Total Factor ⁻
	Total Disease ⁺	Total Disease ⁻	Total

Disease⁺ - Present disease, Disease⁻ - Absent disease, Factor⁺ - present factor, Factor⁻ - absent factor

! Calculations only in the above-the-line or below-the-line component

Statistics

- Risk of illness in those exposed
- $RIE = p(B^+ | F^+) = \frac{a}{a+b}$

B – disease, F - factor

	B⁺	B⁻	
F⁺	a	b	a+b
F⁻	c	d	c+d
	Total B⁺	Total B⁻	Total=n

- that is, the percentage (we can multiply by 100) of sick people among those exposed

Statistics

- Risk of disease in those not exposed
- $RIN = p(B^+ | F^-) = \frac{c}{c+d}$

B – disease, F - factor

	B⁺	B⁻	
F⁺	a	b	a+b
F⁻	c	d	c+d
	Total B⁺	Total B⁻	Total=n

– that is, the percentage (we can multiply by 100) of sick people among those not exposed

Statistics for quantifying the factor-disease link

- Attributable risk:

$$AR = RIE - RIN = \frac{a}{a+b} - \frac{c}{c+d}$$

The difference in illness between those exposed and those not exposed

B – disease, F - factor

	B⁺	B⁻	
F⁺	a	b	a+b
F⁻	c	d	c+d
	Total B⁺	Total B⁻	Total=n

Indicators for quantifying the factor-disease link

- Relative risk:

$$RR = \frac{RIE}{RIN} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

How many times more individuals exposed to the risk factor get sick than those not exposed

Interpretation:

RR = 1 → no association

RR > 1 → increased risk

RR < 1 → protective effect

B – disease, F - factor

	B ⁺	B ⁻	
F ⁺	a	b	a+b
F ⁻	c	d	c+d
	Total B ⁺	Total B ⁻	Total=n

1000 subjects aged 60.

Risk factor: obesity

Disease: osteoporosis

	Osteoporosis ⁺	Osteoporosis ⁻	Total
Obesity ⁺	40	110	150
Obesity ⁻	60	790	850
Total	100	900	1000

Of the 1000 included in the study, 150 were obese. After a 10-year follow-up, 40 of the obese had osteoporosis, 60 of the non-obese had osteoporosis.

Example

		Osteoporosis ⁺	Osteoporosis ⁻	Total
Expose	Obesity ⁺	40	110	150
Non-expose	Obesity ⁻	60	790	850
	Total	100	900	1000

1000 subjects aged 60.

Group 1: with obesity

Group 2: without obesity

at the start of the study none of the patients had osteoporosis

They were followed for 10 years

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!
- it does not result from the study which is the cause and which is the effect

	Osteoporosis ⁺	Osteoporosis ⁻	Total	
Obesity ⁺	40	110	150	26.7
Obesity ⁻	60	790	850	
Total	100	900	1000	

- Risk of illness in those exposed

$$RIE = \frac{a}{a + b} = \frac{40}{150} = 0.267$$

$$RIE = 26,7\%$$

Percentage of illness among those exposed (with risk factor) = 26.7%

	Osteoporosis ⁺	Osteoporosis ⁻	Total	
Obesity ⁺	40	110	150	26.7%
Obesity ⁻	60	790	850	7%
Total	100	900	1000	

- Risk of disease in those not exposed

$$RIN = \frac{c}{c + d} = \frac{60}{850} = 0.07$$

$$RIN = 7\%$$

Percentage of illness among those not exposed (without risk factor) = 7%

	Osteoporosis ⁺	Osteoporosis ⁻	Total	
Obesity ⁺	40	110	150	26.7%
Obesity ⁻	60	790	850	7%
Total	100	900	1000	

- Attributable risk:

$$AR = RIE - RIN = \frac{a}{a + b} - \frac{c}{c + d} = 0.267 - 0.07 = 0.197$$

$$AR = 19.7\%$$

The difference in illness between those exposed and those not exposed was 19.7%

19.7% more obese individuals developed osteoporosis than those who were not obese

	Osteoporosis ⁺	Osteoporosis ⁻	Total	
Obesity ⁺	40	110	150	26.7%
Obesity ⁻	60	790	850	7%
Total	100	900	1000	

- Relative risk :

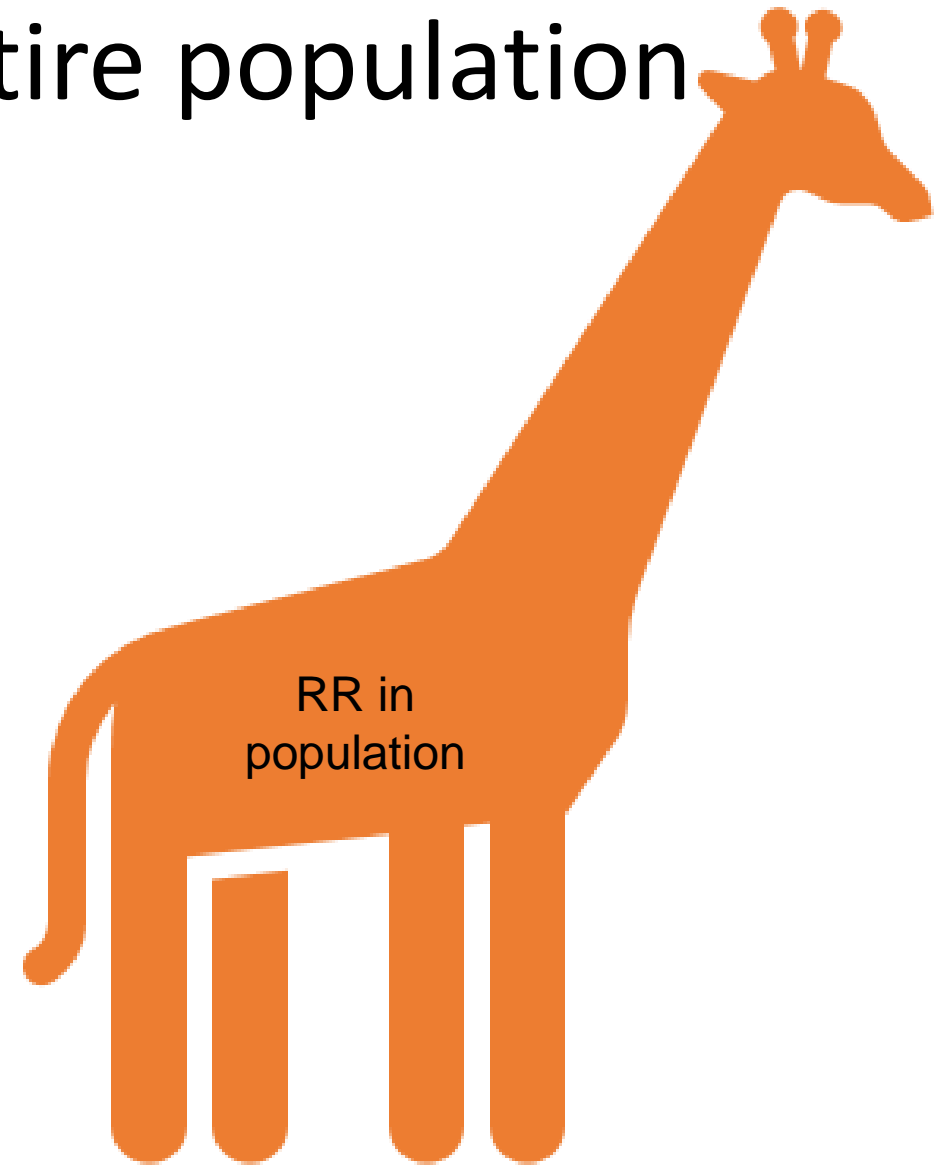
$$RR = \frac{RIE}{RIN} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{0.267}{0.07} = 3.81$$

$$RR = 3.81$$

Obese individuals have a 3.81 times higher risk of osteoporosis than those who are not obese

Generalization to the entire population

- Confidence interval of 95%
- Ex. RIE = 26.7%; 95% IC (9 – 42)
- Ex. RIN = 7%; 95% IC (0.8 – 12)
- Ex. AR = 19.7%; 95% IC (11 – 27)
- Ex. RR = 3.81; 95% IC (3 – 7)



Generalization to the entire population

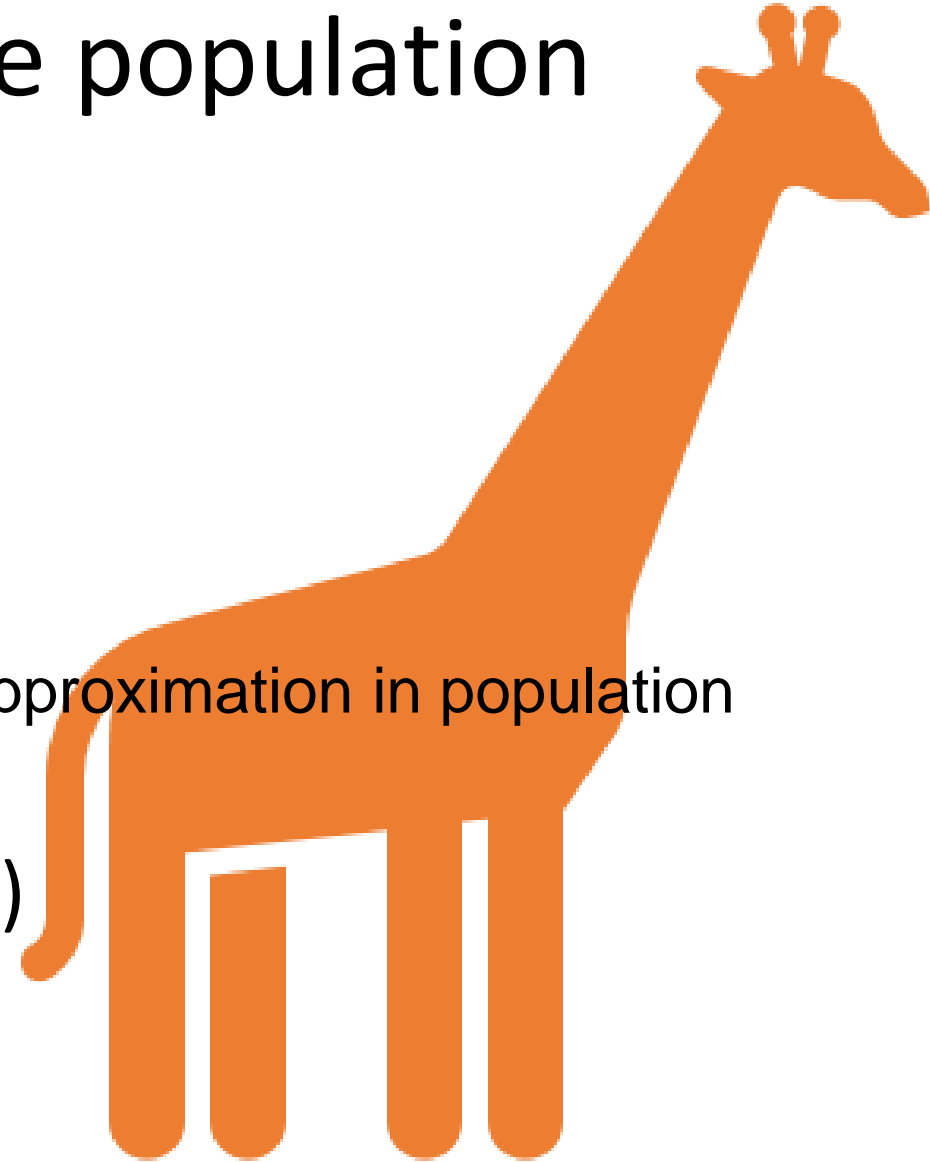
- 95% confidence interval

Ex. RR = 3.81; 95% CI (3 – 7)

calculated from the sample

RA = 19.7%; 95% CI (11 – 27)

approximation in population

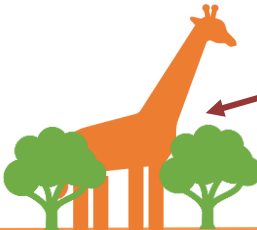


95% confidence interval for RR

calculated on
sample =



RR



!!! We don't know where it is,
it could be anywhere in
between.

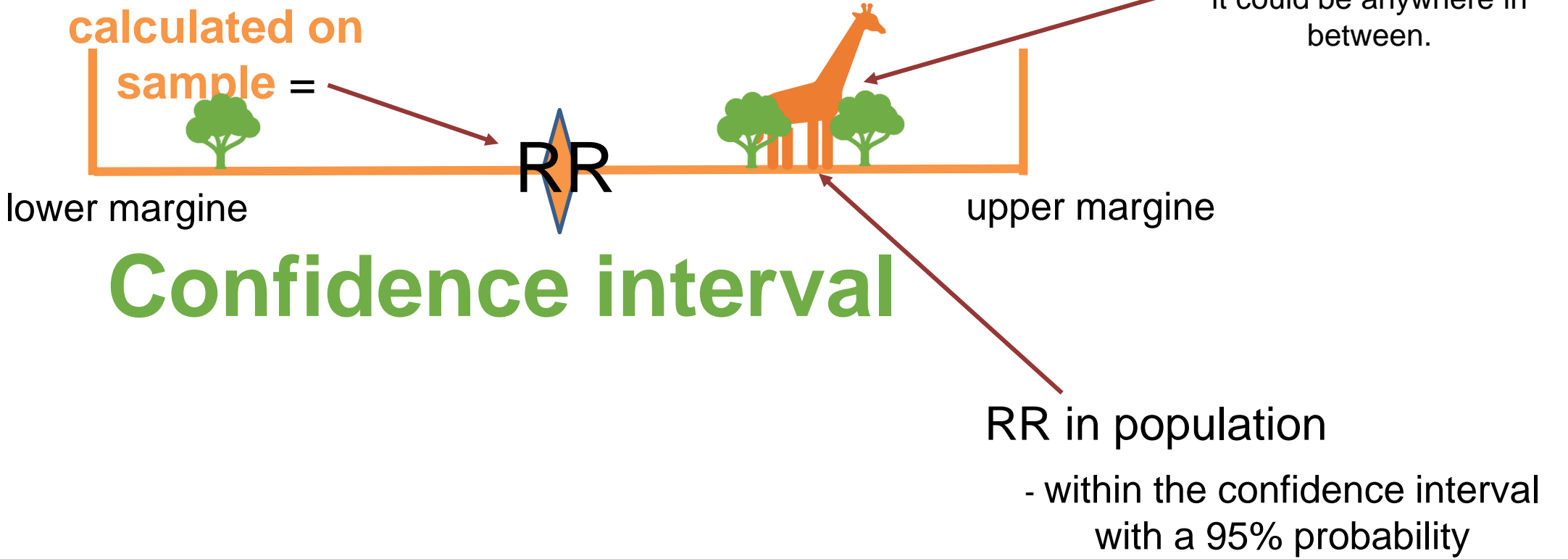
lower margine

upper margine

Confidence interval

RR in population

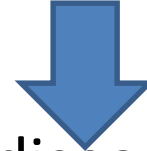
- within the confidence interval
with a 95% probability



Confidence interval interpretation

Do we have a statistically significant association?

statistically **not significant** ($p \geq 0.05$) \leftrightarrow **1** \in 95% CI (value 1 is in the interval)



it is likely that $RR=1 \rightarrow$ the factor and the disease are not associated

statistically **significant** ($p < 0.05$) \leftrightarrow **1** \notin 95% CI (value 1 is not in the range)



it is likely that $RR \neq 1 \rightarrow$ the factor and the disease are associated

!!! $RR = 1$ – there is no relationship

95% CI clinical interpretation

- wide range
 - (4 – 22)
 - imprecise study
- narrow range
 - (3 – 4)
 - precise study
- lower limit much greater than 1
 - (3.6 – 12)
 - important risk factor
- upper limit much less than 1
 - (0.1 – 0.6)
 - important protective factor

95% CI clinical interpretation



- both margins close to 1
 - (1.3 – 1.6)
 - risk factor of little clinical importance

- one edge close to 1, the other far from 1
 - (1.2 – 12)
 - risk factor of unclear clinical importance

- both margins far from 1
 - (10 – 12)
 - risk factor high clinical importance



- both margins close to 1
 - (0.8 – 0.95)
 - protective factor of little clinical importance

- one edge close to 1, the other close to 0
 - (0.2 – 0.95)
 - protective factor of unclear clinical importance

- both margins close to 0
 - (0.2 – 0.3)
 - protective factor of high clinical importance

Cohort study advantages

- cause of the disease
 - we can see that the factor **precedes** the disease
 - in case-control studies we do not know this
- compared to case-control/cross-sectional studies
 - **multiple** risk factors
 - **multiple** endpoints (diseases) can be tracked at the same time

Cohort study advantages

- Allows for the study of **rare** exposure
- **Deaths** due to exposure are also recorded
- **Relative risk**, incidence can be calculated

Cohort study disadvantages

- Long time:
 - Possible loss of some individuals from the study (they leave the country, they do not want to continue the study, etc.); the longer it is, the more these attrition errors appear
 - change of diagnostic definition
 - change of exposure
 - change of investigators
- Very high costs/difficult to organize
- Not useful in the study of rare pathologies (requires huge cohort)
- Possible confounding factors

Bias in Cohort Studies

- Common biases:
 - **Selection bias**
 - **Information bias**
 - **Loss to follow-up bias**
- These can affect validity of results.

Confounding

- A **confounder** is a variable that:
 - is associated with exposure
 - influences the outcome
- Example:
 - Smoking and periodontitis.
 - Possible confounder:
 - **Age**
 - Older individuals may both smoke more and have more periodontal disease.

Methods to Control Confounding

- During design:
 - Restriction
 - Matching
- During analysis:
 - Stratification
 - Multivariable regression

Multivariate analysis

logistic regression

- multiple factors can be analyzed at the same time
- the influence of factors such as gender, age, etc. can be “controlled” or evaluated
 - e.g. we want to see if fractures occur due to calcium deficiency or due to age - it is called adjusting/controlling for age
- the OR and confidence interval will be calculated for each factor
- the interpretation is done analogously to the interpretation of OR in the case of univariate analysis

Loss to Follow-Up

- Participants may drop out during long studies.
- Problems:
 - reduced statistical power
 - potential bias
- Example:
 - Patients moving away from dental clinics.

Cohort vs Case-Control Studies

	Case-control	Cohort
Direction	Disease → Exposure	Exposure → Disease
Indicator	OR	RR
Best for	rare disease	rare exposure
Time	cross-sectional or longitudinal retrospective	longitudinal prospective or retrospective
Groups	Groups divided by the disease	Groups divided by the exposure
Causality	cannot be establish	high probability of causality

Key Take-Home Messages

- Cohort studies follow exposed and non-exposed groups over time
- Allow calculation of incidence and relative risk
- Useful for studying risk factors
- Important design in evidence-based dentistry

Example Cohort study

Scenario - Cohort study

1000 individuals with cardiovascular disease were surveyed about medication adherence (taking prescribed medications).

They were followed for 10 years

Mortality was recorded

Granger BB. Self-reported medication adherence for heart failure is associated with lower risk of all-cause hospitalisation and death. *Evid Based Nurs.* 2015 Oct;18(4):123.

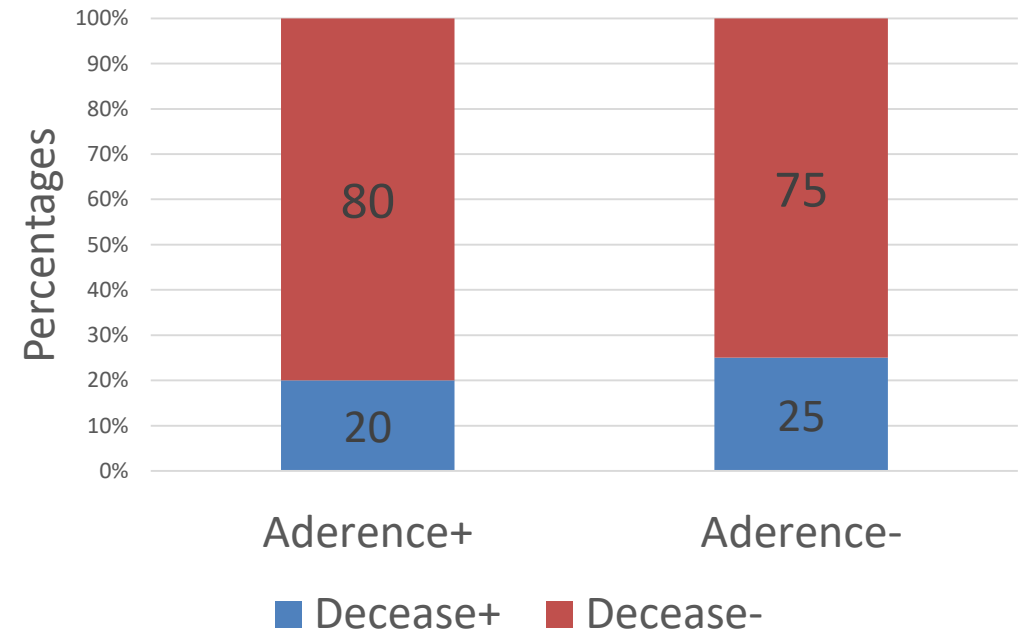
Scenario

- Results:
- Of the 1000 patients, 600 declared adherence to medication
- Of these, 120 died within the next 10 years
- Of the remaining 100 died within the next 10 years

Observed contingency table

	Decease ⁺	Decease ⁻	Total
Aderence ⁺	120	480	600
Aderence ⁻	100	300	400
Total	200	800	1000

- $RIE=120/600=0.20$
 - 20% of patients with medication adherence died in the next 10 years
- $RIN=100/400=0.25$
 - 25% of patients without medication adherence died in the next 10 years
- We will compare 20% with 25%



(! fictive data)

Chi-square test

Null hypothesis (H0):

- Stated adherence to medication and mortality in patients with cardiovascular disease (CVD) are independent

Alternative hypothesis (H1):

- Stated adherence to medication and mortality in patients with CVD are dependent

Results

- $p=0.061$
- $p>0.05$ we failed to reject H0:
- Stated adherence to medication and mortality in patients with CVD are independent

RR – Relative risk

- $RR = \frac{RIE}{RIN} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{0.20}{0.25} = 0.8$
- $RR > 1$ indicates a risk factor

95% confidence interval (CI) for RR

- RR=0.8,

95% CI 0.6-1.2

- did not indicate the presence of a prognostic factor
 - value 1 is in the range, RR in the population can be equal 1, which indicates that there is no risk factor
- interval is narrow
 - so the study is precise

AR – attributable risk

- $AR = RIE - RIN = \frac{a}{a+b} - \frac{c}{c+d} = 0.20 - 0.25 = -0.05$
- AR = -5%
- 95% CI (-9; 2)
- 0 is in 95% CI → there is no association between the studied factor and the disease

!

- If the statistical test is not significant
 - Do not interpret the result clinically
- Calculating the RR, AR and confidence interval is redundant and does not provide new information

Example of articles

Serum Levels of 25-Hydroxyvitamin D in Patients with Seborrheic Dermatitis: A Case-Control Study

Siavash Rahimi , Negar Nemati , Seyedeh Sareh Shafaei-Tonekaboni

Abstract

This study was aimed at comparing the 25-hydroxyvitamin D (25(OH)D) status in patients with facial or scalp seborrheic dermatitis with healthy subjects.

This **case-control study** included 289 patients (118 with Seborrheic Dermatitis and 171 **sex- and age-matched control** subjects) from the outpatient clinic of two hospital dermatology departments in the west of Mazandaran province, Iran.

Serum mean \pm standard deviation of 25(OH)D levels were significantly lower in seborrheic dermatitis patients than in control subjects (20.71 ± 8.16 vs. 23.91 ± 7.78 , $P = 0.007$). **Vitamin D under 30 ng/ml was associated with OR: 4.22 (95% CI: 1.077-16.534, $P = 0.039$) for seborrheic dermatitis.** The severity of scalp disease was significantly associated with serum 25(OH)D level ($P = 0.003$). Cases with severe scalp scores had significantly lower serum 25(OH)D level compared to moderate OR score ($P = 0.036$). A similar trend was not seen in the facial disease. The 25(OH)D values are significantly lower in seborrheic dermatitis patients than in healthy subjects. Furthermore, the scalp disease severity was associated with lower serum 25(OH)D level.

Our results may suggest that vitamin D may play a role in the pathogenesis of seborrheic dermatitis.

OR=4.22 (95% CI: 1.077-16.534, $P = 0.039$, patients with Vitamin D deficiency had a 4.22-fold higher risk of seborrheic dermatitis than those with normal vitamin D levels. Note the concordance between the lack of value 1 in the confidence interval and $p < 0.05$. However, only age and sex were controlled, not the diet deficient in Calcium. The study is quite imprecise, the confidence interval is very large (due to the small number of patients included in the study).

Association between coffee and green tea intake and pneumonia among the Japanese elderly: a case-control study

Kyoko Kondo 1, Kanzo Suzuki 2 3, Masakazu Washio 4, Satoko Ohfuji 5 6, Satoru Adachi 7, Sakae Kan 8, Seiichiro Imai 9 10, Kunihiro Yoshimura 11, Naoyuki Miyashita 12, Nobumitsu Fujisawa 13, Akiko Maeda 5, Wakaba Fukushima 5 6, Yoshio Hirota 14 15, Pneumonia in Elderly People Study Group

Abstract

A large prospective cohort study in the United States examined the association between coffee intake and overall and cause-specific mortality and showed an inverse association between pneumonia and influenza deaths and coffee intake. In Japan, the mortality rate of pneumonia in elderly people is high, and its prevention is an important issue. The present study investigated the association between coffee and green tea intake and pneumonia among the elderly.

The design was a hospital-based **case control study**. The **cases** were patients over 65 years old newly diagnosed as pneumonia. As a **control**, **patients with the same sex and age** (range of 5 years) who visited the same medical institution around the same time (within 2 months after examination of the case) for a disease other than pneumonia were selected. **There were two controls per case**. Odds ratio (OR) and 95% confidence interval (CI) for pneumonia of coffee and green tea intake during the past month were calculated using a conditional logistic regression model.

A total of 199 cases and 374 controls were enrolled. When compared to those who do not drink coffee, the **OR for pneumonia of those who drink less than one cup of coffee per day was 0.69 (95% CI 0.39-1.21)**, **OR of those who drink one cup was 0.67 (0.38-1.18)**, and **OR of those who drink two or more cups was 0.50 (0.28-0.88) (Trend p = 0.024)**. No association was found between pneumonia and green tea consumption. This study suggested a preventive association between coffee intake over 2 cups per day and pneumonia in the elderly.

OR=0.69 (95% CI 0.39-1.21) and OR=0.67 (0.38-1.18) so drinking one cup of coffee or none is not a protective factor for pneumonia because 1 is in the interval, OR=0.50 (0.28-0.88) drinking two cups of coffee or more is a protective factor for pneumonia because 1 is not in the interval. Note the agreement between the lack of value 1 in the confidence interval and $p < 0.05$. However, only age and sex were controlled, cardiovascular diseases were not.

Obesity during the COVID-19 pandemic: both cause of high risk and potential effect of lockdown? A population-based electronic health record study

M. Katsoulis,^{a,b} L. Pasea,^{a,b} A.G. Lai,^{a,b} R.J.B. Dobson,^{a,b,c} S. Denaxas,^{a,b} H. Hemingway,^{a,b} and A. Banerjee^{a,b,d,e,*}

Objectives

Obesity is a modifiable risk factor for coronavirus disease 2019 (COVID-19)–related mortality. We estimated excess mortality in obesity, both ‘direct’, through infection, and ‘indirect’, through changes in health care, and also due to potential increasing obesity during lockdown.

Study design

The study design of this study is a **cohort study** and causal inference methods.

Methods

In population-based electronic health records for 1,958,638 individuals in England, we estimated 1-year mortality risk (‘direct’ and ‘indirect’ effects) for obese individuals, incorporating (i) pre-COVID-19 risk by age, sex and comorbidities, (ii) population infection rate and (iii) relative impact on mortality (relative risk [RR]: 1.2, 1.5, 2.0 and 3.0).

Results

For severely obese individuals (3.5% at baseline), at 10% population infection rate, we estimated direct impact of 240 and 479 excess deaths in England at **RR 1.5**. Owing to BMI change during the lockdown, we estimated that 97,755 (5.4%: normal weight to overweight, 5.0%: overweight to obese and 1.3%: obese to severely obese) to 434,104 individuals (15%: normal weight to overweight, 15%: overweight to obese and 6%: obese to severely obese) would be at higher risk for COVID-19 over one year.

Conclusions

Prevention of obesity and promotion of physical activity are at least as important as physical isolation of severely obese individuals during the pandemic.

Practice and outcomes of airway management in patients with cervical orthoses

Yu-Chun Liu, Wen-Cheng Huang, Elise Chia-Hui Tan, Shiang-Suo Huang, Yen-Kai Wang, Ya-Chun Chu
<https://doi.org/10.1016/j.jfma.2021.02.002>

Increasing evidence indicates an association of video laryngoscopy with the success rate of airway management in patients with neck immobilization. Nevertheless, clinical practice protocols for tracheal intubation in patients immobilized using various types of cervical orthoses and the outcomes remain unclear.

Methods

We **retrospectively** assessed the tracheal intubation techniques selected for patients immobilized using cervical orthoses from 2015 to 2018. The endpoints were **the intubation outcomes** of the different techniques and the factors associated with the selection of the technique.

Results

We included 218 patients, 118 of whom wore **halo vest braces** (halo vest group) and 100 wore **cervical collars** (collar group). GlideScope video laryngoscopy (GVL) and fiberoptic bronchoscopy (FOB) were the initial intubation methods in 98 and 120 patients, respectively. GVL had a higher first-attempt success rate than did FOB in the collar group ($p = 0.002$) but not in the halo vest group ($p = 0.522$). **GVL was associated with a lower risk of episodes of $\text{SaO}_2 < 90\%$ (adjusted relative risk [aRR], 0.11; 95% CI, 0.02–0.67; $p = 0.016$).** However, in the halo vest group, more frequent requirement of a rescue technique ($p = 0.002$) and necessity of patient awakening ($p = 0.001$) was noted when GVL was used. Use of the halo vest brace and noting of severe cord compression were independent predictors of the initial selection of FOB.

Conclusion

Caution should be exercised when using GVL for tracheal intubation in patients immobilized using halo vest braces.

The risk of low oxygen saturation was approximately 10 times lower ($\text{RR} < 1$) in the group wearing the collar-type cervical brace than in those wearing the Halo-type adjustable cervical brace. The risk was adjusted for intubation method to control for the effect of this factor.

Assessment of dental caries predictors in 6-year-old school children - results from 5-year retrospective cohort study.

Masood M1, Yusof N, Hassan MI, Jaafar N.

BACKGROUND: This was a **retrospective cohort** study undertaken to assess the rate and pattern of dental caries development in 6-year-old school children followed-up for a period of 5 years, and to identify baseline risk factors that were associated with 5 years caries.

METHODS: This 5-years retrospective cohort study comprised primary school children initially aged 6 years in 2004. Caries experience of each child was recorded annually using World Health Organization criteria.

RESULTS: The sample consisted of 1830 school children. All components of caries showed significant differences between baseline and final examination. Results revealed the **initial baseline caries level in permanent dentition was a strong predictor for future caries after 5 years (RR=3.78, 95% CI=3.48-4.10, p<0.001)**. Logistic regression analysis showed significant association between caries occurrence and residence (urban) (OR=1.80, p<0.001). However, it was not significantly associated with gender and ethnicity.

CONCLUSION: The majority of 12-year-old school children (70%) were caries-free and most of the caries were concentrated in only a small proportion (30%) of them. We found that the presence of caries in permanent teeth at the age of 6 years was a strong predictor of future caries development in this population.

Thank you!