Risk and causality

Assoc. Prof. Dr. Emilian Damian Popovici
Univ. Assistant Dr. Luminiţa Bădiţoiu

Plan of presentation

- **Risk Factor - Definitions, Types, Risk Factor – Disease Relationship ;**
- **Types of studies used in risk analysis;**
- **Errors in such studies;**
- **Errors control;**
- **Causality - Hill Causality Guide.**

Risk factor

- a behavioral, genetic, demographic, environmental or physiological aspect that increases the likelihood of a healthy person's illness.

There may be different **types** :

- **Genetics -** Family hypercholesterolemia;
- **Environmental -** toxins, pollutants, drugs;
- **Social -** divorce, unemployment, death of a family member;
- **Behavioral -** smoking.

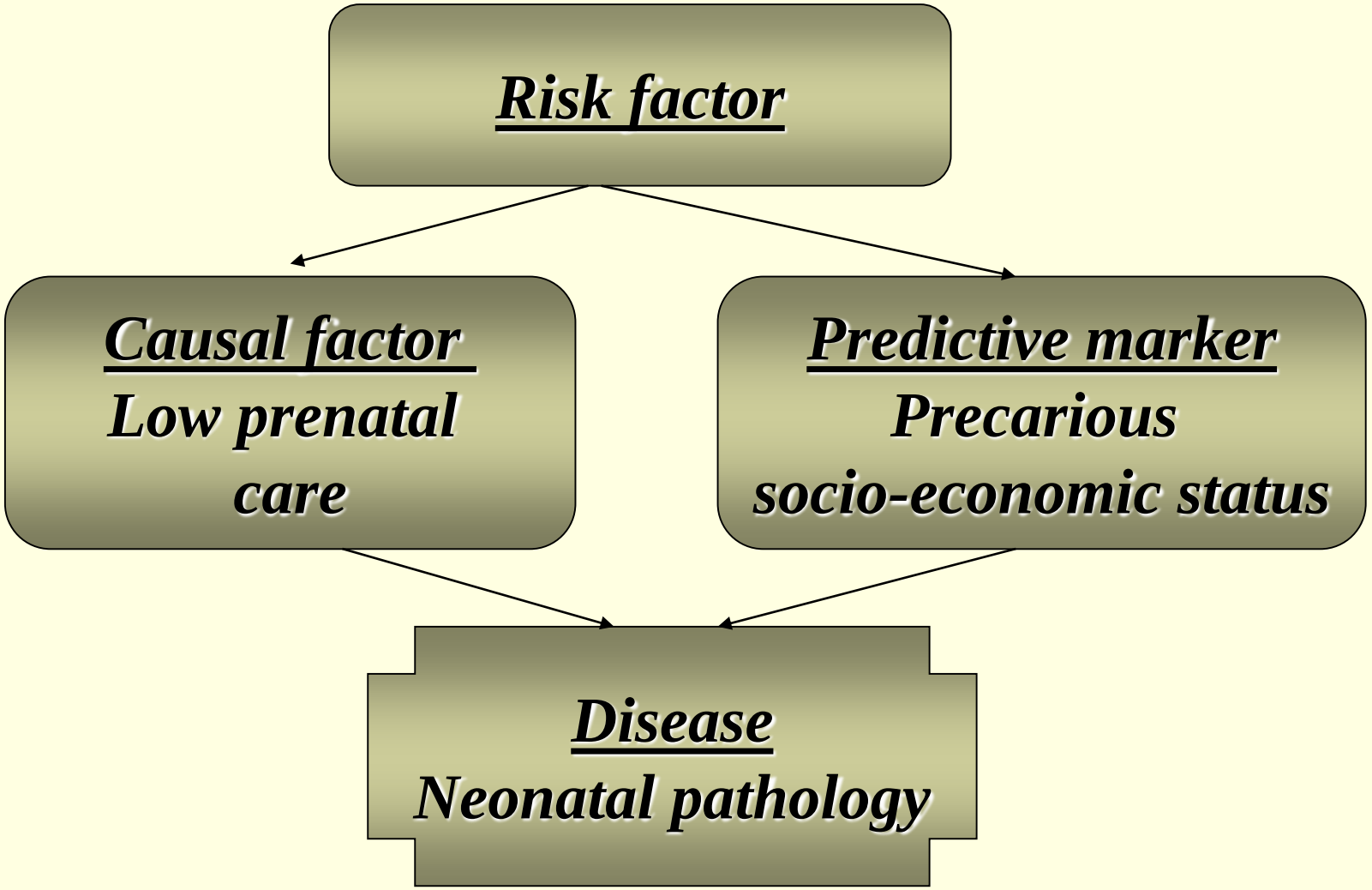
Exposure

- contact with / manifestation of a risk factor;

- It can be:
  - **Single, accidental -** a traffic accident can expose the entire community to a toxic substance;

  - **Chronic -** sedentarism, drug abuse;
    - the measurement of such an exposure shall include:
      - ✓The current dose,
      - ✓The cumulative dose,
      - ✓Years of exposure.

Risk factor / disease relationship



Risk analysis

1. **Experimental studies:**
   - controls and manipulates the risk factor;
   - They're more scientifically rigorous;
   - Ethical reasons limit their use in the study of risk factors.

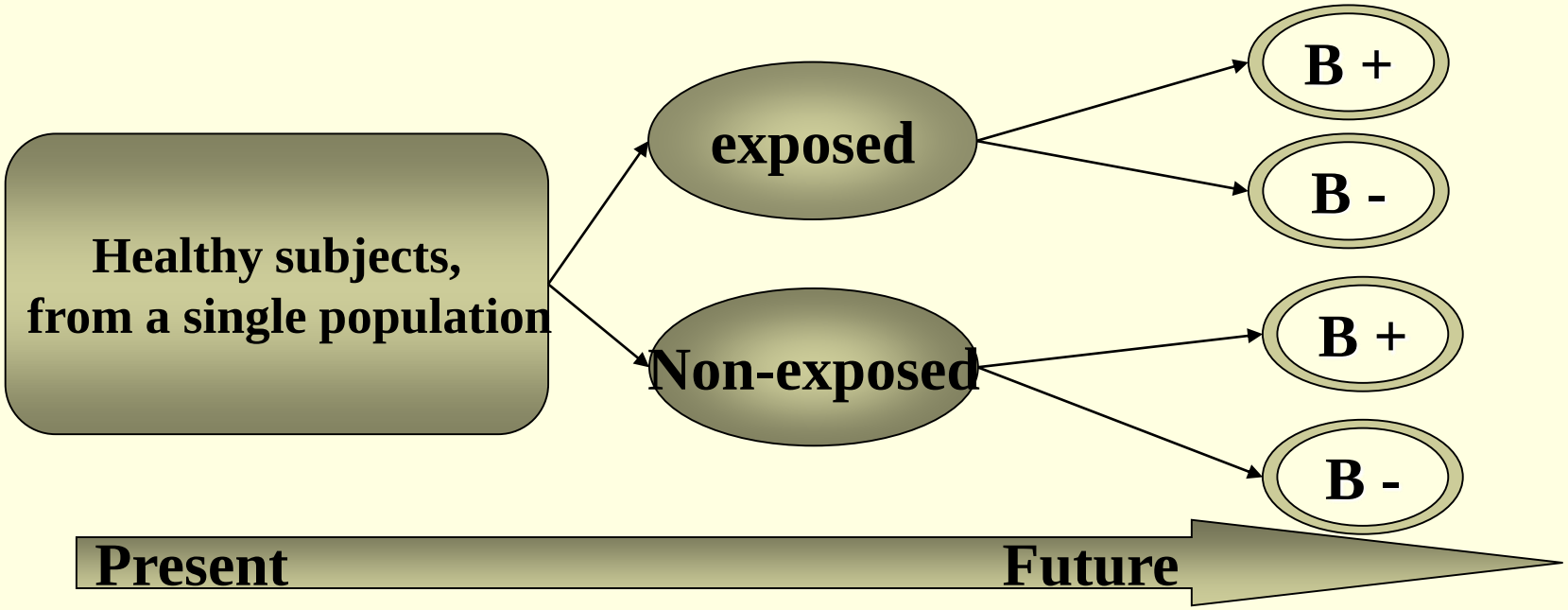
2. **Observational studies:**
   - Without the investigator's active intervention;
   - Sometimes they are the only ones feasible to evaluate the risk - disease association.

Observational studies

- Can be:
  - **Descriptive studies** – estimates the frequency of the disease and pays attention to the potential risk-sickness factor association;
  - **Analytical Studies -** attempts to explain how the disease is distributed by defining the risk factor and establishing the cause-effect link;
- There is no control over the experiment, causing more errors compared to experimental studies!

1. Prospective studies

- Synonym terms: cohort studies, follow-up study,
- The sample of *healthy* subjects exposed to suspected risk factor action is tracked in time from the start of the study.



1. Prospective studies

Contingency table 2 x 2

	B +	B -	Total
E +	a	b	a+b
E -	c	d	c+d
Total	a+c	b+d	

- A significantly higher incidence of disease in the group of exposed reveals the association between risk factor and disease,
- A significantly lower incidence of disease in the group of exposed reveals the protective effect.

1. Prospective studies

- The power of the association between the risk factor and the disease is shown by :
  - **RR =** risk of disease in the presence of the risk factor / risk of disease in the absence of the risk factor

$$RR = \frac{P(B+/E+)}{P(B+/E-)} = \frac{a/a+b}{c/c+d}$$

- **Odd ratio =** the likelihood of exposed people to get sick / the likelihood of unexposed people to get sick

$$OR = \frac{P(B+/E+)}{P(B-/E+)} \Big/ \frac{P(B+/E-)}{P(B-/E+)} = \frac{a/b}{c/d}$$

1. Prospective studies

Usually OR >RR

OR > 1 (RR >1)
The entire confidence interval > 1

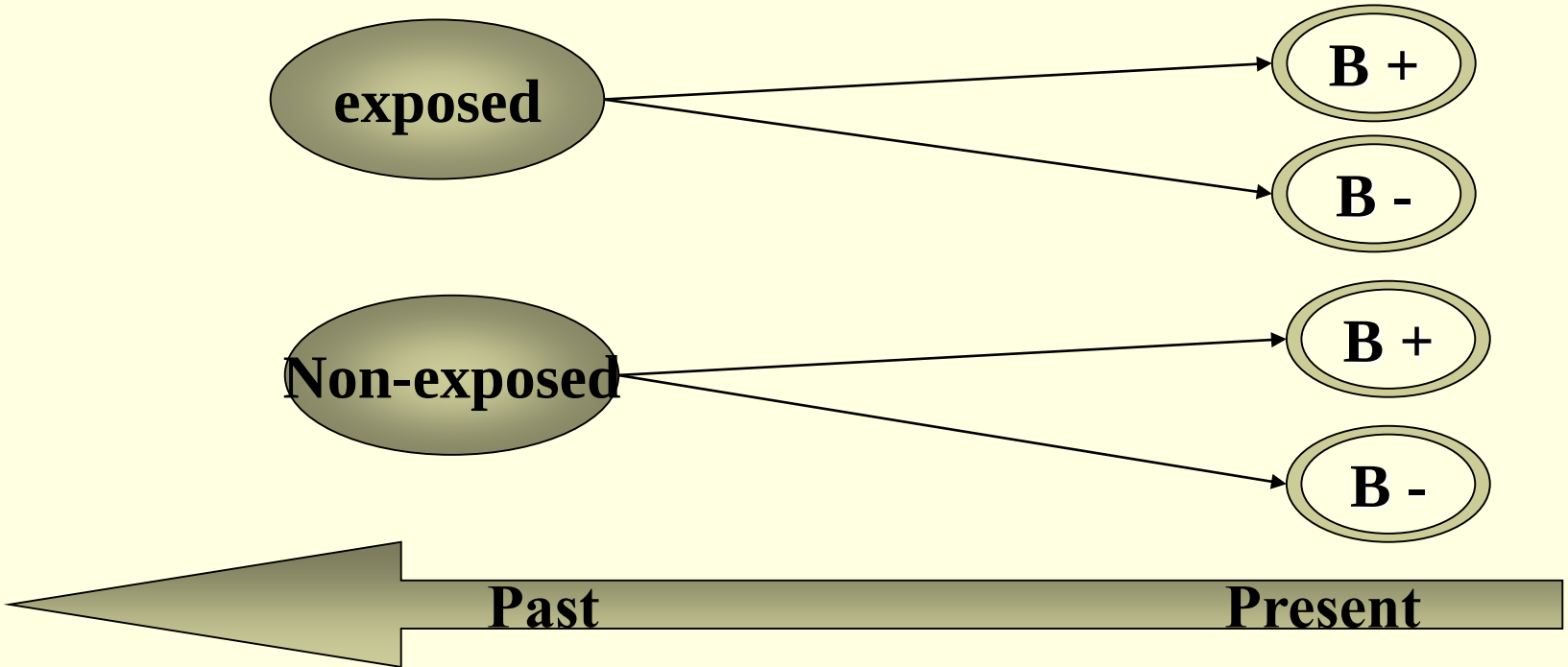
Risk factor!
No matter the type of study

1. Prospective studies

- **Advantages:**
  - The only observational studies that allow direct risk estimation, meaning the probability of the healthy subject to develop the disease within a specified time period;
  - The closest observational study to the experimental one;
  - It is possible to simultaneously examine the association between the hypothetical risk factor and several disorders.
- **Disadvantages :**
  - For rare diseases, it is necessary to supervise an extended sample over a long period of time;
  - Exposure to a risk factor raises ethical issues.

2. Retrospective cohort studies

- Synonyms: historical cohort studies;
- The cohort of exposed and non-exposed subjects is tracked retrospectively, based on previous records.

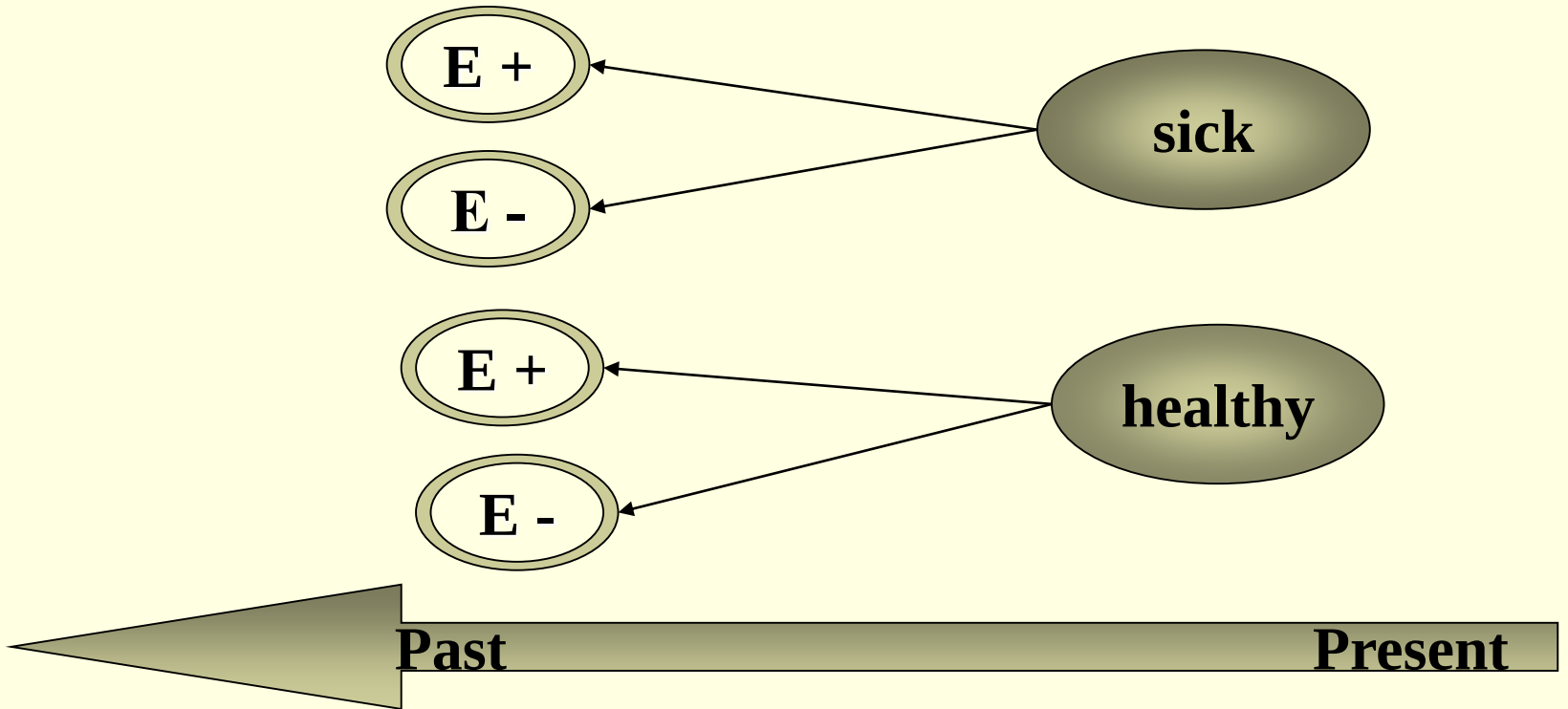


2. Retrospective cohort studies

- **Advantages:**
  - They can be done quickly and at minimal cost;
- **Disadvantages :**
  - Depends on existing medical records;
  - If these are incomplete or inaccurate, the findings of the study may be erroneous.

3. Case-control studies

- Synonyms: case-referent studies,
- Subjects classified as sick and healthy (at the start of the study) are followed up retrospectively for the determination of exposure to a hypothetical risk factor.



3. Case-control studies

- **Selection of samples :**

  - Errors are minimal if both cases and control are randomized similar samples from the same population;
  - Both cases and control are selected according to the same criteria, especially when the group of cases is not representative of the general population;
  - Due to the difficulty of obtaining comparable groups, multiple control groups are often used :
    - ✓ If similar results are obtained from all control groups, there is a high likelihood that the observation is real;

3. Case-control studies

- **Determination of exposure to the presumed risk factor:**
  - Existing medical records;
  - Interviews;
  - Questionnaires;

- **Measurement of exposure can be done:**
  - **Dichotomous** – smoker / non-smoker;
  - **Polichotomous** – non-smoker, occasional smoker, moderate, inveterate;
  - **Continuous** – consumption of nicotine over a certain period.

3. Case-control studies

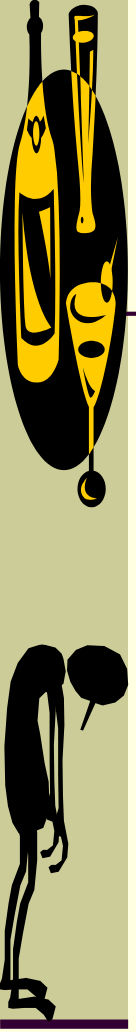
Contingency table 2 x 2

	E+	E -	Total
B +	a	b	a+b
B -	c	d	c+d
Total	a+c	b+d	

- A statistically significant difference between patient exposure and control cases highlights the association between risk factor and disease (chi^2);
- Measurement of association is done through **OR;**
- For continuous variables, correlation and regression are used;

3. Case-control studies

- **Advantages:**
  - Very useful for the study of rare diseases;
  - For situations where between exposure and the onset of the disease a long time passes, but the need to identify cause-to-effect relationships is urgent! (eg, the study of the link between other sexually transmitted diseases and the risk of contracting HIV / AIDS);
  - They can be done quickly and at minimal cost;
- **Disadvantages:**
  - Much more vulnerable to errors;
  - It does not measure the risk directly;
  - It can examine the link to a single disease;
  - It does not determine the temporal relationships between the risk factor and the disease.



3. Case-control studies

Eg: a **case-control** study on the association between depression and alcoholism reveals that patients who have been decontaminated have experienced more frequent depressions in the previous 5 years compared to non-alcoholic subjects.

It can not be established whether alcoholism is secondary to depression or whether depression is the consequence of alcohol!

4. Transversal studies

- Synonyms: Cross-sectional, prevalence study.
- Caught the situation at a time on classification in exposed / non-exposed, sick / healthy;
- Comparing prevalence between exposed and non-exposed;
- The statistical significance of the association between the risk factor and the disease is determined by chi2;
- The size of the association is established by **OR**.

4. Transversal studies

- **Advantages:**
  - They can be the basis for a future prospective study;
  - They may be useful for defining samples of a case-control study;
  - They can be done quickly and at minimal costs;

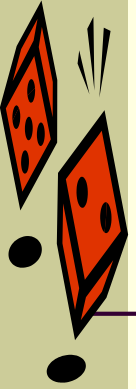
- **Disadvantages:**
  - It presents the same limits as in case-control studies;
  - Prevalence may be tainted by excluding cases of deceased or healed rapidly;

Errors in risk analysis observational studies

- **Systematic differences** between the groups compared may affect internal validity and invalidate the findings of the study;

- **Variables may be confused or there may be other interventions** that are not taken into account;

- **Eg**: smoking is usually associated with coffee consumption and this association may create confusion in vasoconstriction studies.



Errors in risk analysis observational studies

Eg: A prospective cohort study looked at the incidence of neoplasia over a 10-year period on two randomized samples, one in the state of Nevada, where gambling is legal, the other one in Utah, where they are not legal. The high incidence of oncological pathology in the first batch may lead to the conclusion that gambling is a risk factor ?!!

Error of confusion: excessive drinking and smoking among Nevada subjects was mistaken for participation in gambling!

The control group in Utah, most of them being Mormons, was not made up of smokers or alcohol consumers.

Errors in risk analysis observational studies

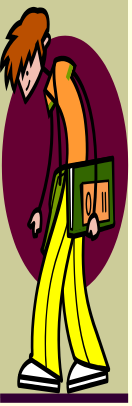
- **Selection Errors :**

  - The investigator compares groups of subjects, also different in other elements than the disease state or risk factor!

  - **Eg:** a study on the effect of jogging on coronary artery disease compares the incidence of disease among jogging practicants versus a sample of the general population.
  - **Selection error** – people who carry out physical maintenance are generally more caring about their health, have a low-in-fat diet - which can affect any conclusion of the study!

Errors in risk analysis observational studies

- **Migration errors :**
  - Appears when the subjects are lost from the study or move from one group to another;
  - **Eg.** A prospective study on the effects of nutrition on school performance, performed on a group of students in a particular school, considered having a good nutrition and a group of children in a neighborhood school, considered having a worse nutrition. Losing some of the subjects in the second group as a result of dropping out of the education process at a higher rate than the first sample will invalidate the results of the study!

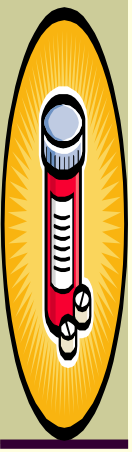


Errors in risk analysis observational studies

- **Measurement / surveillance errors** occur as a result of systematic differences in the measurement of the variable among the two groups;

- **Eg:** a prospective study to determine the association between postmenopausal estrogen administration and the development of uterine neoplasms. Because the treated group is gynecologically controlled more often, more cases may be identified even if the incidence is similar in both samples!

Errors in risk analysis observational studies



Errors of information:

- Patients and their visitors are more persistent in identifying past exposures versus the control group. Sometimes they can overestimate or underestimate exposure systematically.

- **Eg.** mothers of newborn babies with malformations tend to overestimate drug use during pregnancy!

Errors in risk analysis observational studies

- **Sampling errors:** there are systematic differences between the studied sample and the general population, affecting the external validity and therefore the possibility of generalizing the results;

  - Case-control prevalence studies are more susceptible to these biases than those based on incidence because they exclude patients deceased by illness or those who are healing rapidly;

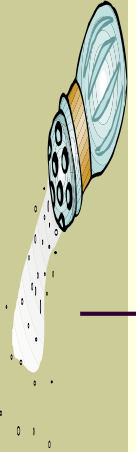
Errors in risk analysis observational studies

- Control groups selected from hospitalized patients generally have a higher morbidity than the general population;

- **Eg. Berkson bias :** case-control study on the association between asthma and emotional disturbances in children. Because children with asthma and psychological disorders are more often hospitalized, the association between asthma and emotional disturbances is artificially increased (unrepresentative outcome for the general pediatric population)

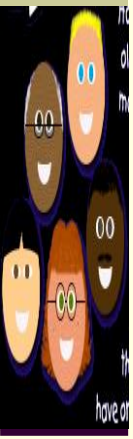
Errors in risk analysis observational studies

- Sampling errors due to **subject motivation** - generating differences between the sample studied and the general population;
- When the subjects lost from the study are **systematically different** from those left, the results can only be applied to the final sample;
- Even though the sample was initially representative of the population, **migration biases** may restrict the generalization of results;
- The more missing subjects, the less generalized are the results!



Error control

1. **Restrict** subject access to the study to minimize confusion :



- **Eg:** Black people consume more salt than other racial groups, so the association between hypertension and salt can be confused with the link between that pathology and race. To avoid confusion, subjects belonging to a single race will be accepted.

- The link between the restricted factor and the disease can not be studied;

Error control

2. **Using pairs of subjects**, similar in many ways;

- **Eg:** In a study on the role of smoking in baldness, cases were selected from men who sought specialized healthcare for this problem and the control group from patients in the family medicine cabinet. Both baldness and smoking are more common in the elderly. In order to eliminate potential confusion, the pairs - a person with baldness / no baldness, were the same age.

Error control

3. **Stratification:**
   - Subgroup of study subjects based on similar features and data analysis for each sub-sample separately;
   - **Eg:** In a case-control study, an association was found between caffeine and coronary artery disease;
   - A detailed analysis shows that smokers consume more coffee than non-smokers and the sample of patients with C.A.D. includes more smokers;
   - To avoid confusion between caffeine and smoking, both the patient group and the control will be stratified according to the smoking status.

Error control

4. **Standardization of rates :**

   From the observational studies you can calculate:

   - **Gross rates** of morbidity / mortality;
   - **Specific rates** - specific mortality by age group;
   - **Adjusted, standardized rates** to compare populations with different fundamental characteristics;

Error control

5. **Assuming "the worst situation":**
   - When confusion is unavoidable or has a minor impact;
   - The effect of confusion is estimated by assuming the "most unfortunate" factor distribution among the compared groups;
   - It is particularly useful in controlling sampling errors due to noncompliance;

   - **Eg:** a study to determine the incidence of coronary artery disease - consideration of all included but non-compliant subjects as pathology cases.

Error control

6. **Statistical methods** - to adjust values of the dependent variable against the influence of one or more variables - sources of confusion;

- **Multivariate regression :**
  - ✓ Logistic regression,
  - ✓ Hazard proportional regression model – Cox;
  - ✓ Analysis of covariance

The Hill Causality Guide

- In order to have a **cause-effect relationship**, the following criteria must be met :

  - **The power of association** - strong association (measured by RR / OR) and statistically significant between a possible cause and a possible effect advocates a cause-effect relationship, more than a weak association;
  - **Consistency** - Several researchers, using different types of study, at different times, circumstances and locations, reach the same conclusions;
  - **The right temporal relationship** - exposure must precede the disease;

The Hill Causality Guide

- **The dose-effect relationship** - the risk is directly proportional to the intensity of the exposure;
- **Reversibility** - Causal association is strengthened when the removal of the cause leads to a decrease in the risk of illness;
- **Plausibility** - a cause-effect association is plausible if it conforms to recognized scientific knowledge; the lack of plausibility may reflect the lack of knowledge rather than the irrevocable absence of causality;

The Hill Causality Guide

- **Specificity** - a single possible cause is related to a single effect (in infectious diseases, genetic diseases);

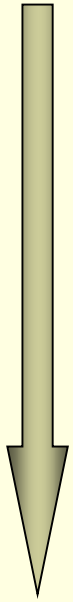
- **Analogy** - the existence of another cause-effect relationship similar to the one studied gives it credibility

Hill criteria are necessary but not sufficient to establish a causal relationship!

The power of studies in determining causality

- **Experimental study**
- **Prospective cohort study**
- **Retrospective cohort study**
- **Case-control study**
- **Transversal study**

The strongest



The weakest

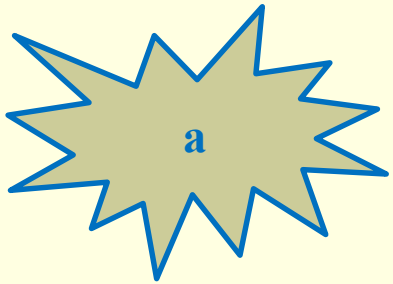
Bibliography

- **Rebecca Knapp, M.Clinton Miller III** – Risk and Causality, in Clinical Epidemiology and Biostatistics,Williams & Wilkins, Baltimore, Maryland, 1992, 109-122

- **G.I.Mihalaş** – Methodology of medical scientific research, Postgraduate course, UMF Timişoara, 2006

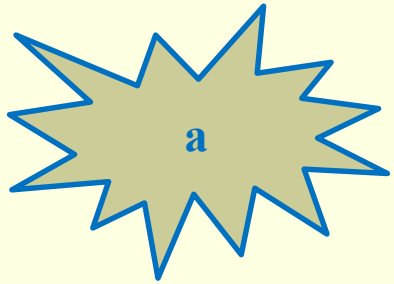
Practical application

- **An epidemiological study described the association between a given exposure and a particular illness as "a strong but statistically insignificant association." Which of the following best describes this :**
  - **a. relative risk = 10,0, p = 0,20**
  - **b. relative risk = 1,3, p = 0,03**
  - **c. relative risk = 1,5, p = 0,01**
  - **d. relative risk = 10,0, p = 0,01**
  - **e. relative risk = 1,5, p = 0,20**



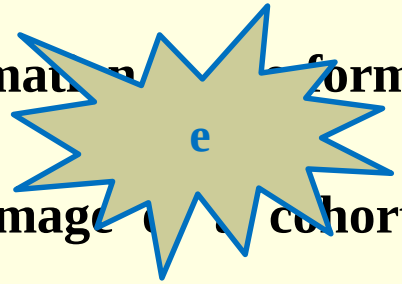
Practical application

- **In a study of 50 cases of disease and 50 controls, it was determined that the difference found in report to a possible etiological factor was not statistically significant. It can be concluded that :**
  - **a. there is no association of the factor with the disease;**
  - **b. the difference may be clinically significant;**
  - **c. the difference may be the result of a variation in the sample;**
  - **d. comparability of cases with witnesses has been confirmed;**
  - **e. interference of the observer or interviewer has been removed.**



Practical application

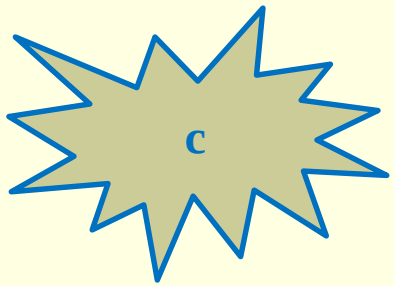
- **A study in England shows that out of 224 families in which a case of polio occurred, in 56 of them there were parrots. In another English study of 99 patients surveyed, 30 had parrots. The conclusion that there is a relationship between the presence of parrots in homes and the appearance of polio in members of that family is :**
  - **a. correct;**
  - **b. incorrect, because it does not specify whether it is about incidence or prevalence;**
  - **c. incorrect, because we are given the information in the form of a frequency, not a risk;**
  - **d. incorrect, because it seems to be the image of a cohort phenomenon;**
  - **e. incorrect, because there is no control or comparison group.**



Practical application

In the family past of 100 schizophrenic patients, 45 of them were found without one of their parents before the age of 18. Of the 134 medical students, 23 were found to have lost one of their parents to the same age. Based on this statistically significant difference, the inference that there is a causal association between the loss of one parent in childhood and schizophrenia is :

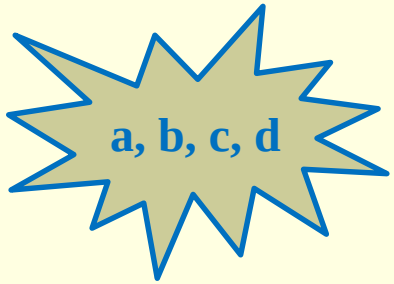
- **a. correct;**
- **b. incorrect, because the comparison is not based on rates;**
- **c. incorrect, because the group of patients and the group of witnesses are not comparable;**
- **d. incorrect, because the observation error was not excluded;**
- **e. none of the above.**



Practical application

To study whether there is a relationship between the use of oral contraceptives (estro-progestative) and cervical cancer, a group of women between 18 and 58 years of age, without colorectal cancer at the beginning of the study, was followed from December 1976 to January 1980. Surveillance consisted mainly of the annual Papanicolau smear, investigating the use of contraceptives in a questionnaire. We're dealing with an investigation :

- **a. longitudinal;**
- **b. prospective;**
- **c. of the cohort;**
- **d. analytical;**
- **e. case-control.**

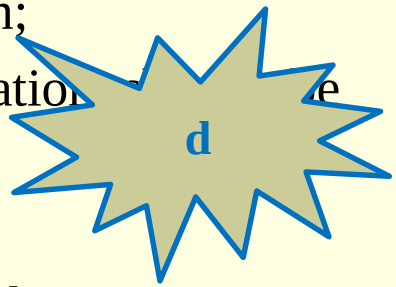


Practical application

- The following table lists the results of the survey depending on the duration of use of contraceptives :

Exposure duration	They did not take contraceptives	1-4 years	>4 years
The incidence of cervical cancer(from 100 000)	32	96	…?
Relative risk	1	3	5

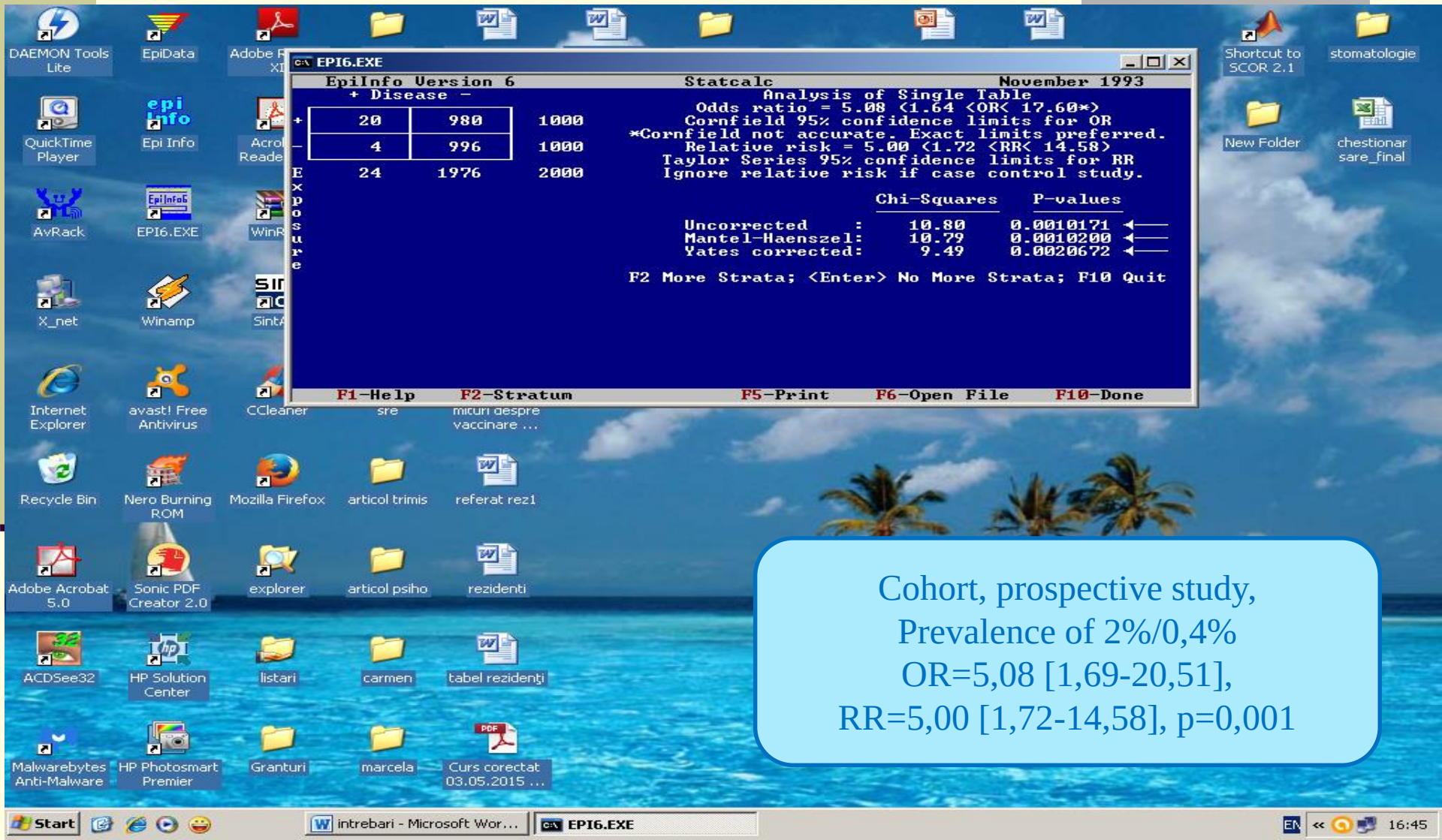
- Do you have enough elements to complete the table and calculate the incidence of cervical cancer in women using contraceptives for more than 4 years?
  - a. no, the number of women who have been exposed for more than 4 years should be known;
  - b. no, because the incidence of non-exposed is unknown;
  - c. no, the incidence of cancer in the general population should be known;
  - d. yes, the incidence is 160/100000;
  - e. no, you should know the number of non-exposed smokers.



Practical application

- **In the year 1955, 1000 women working in a radium painting factory of clock dials and 1000 telephone workers were included in a study. Every new cases of bone cancer were recorded up to 1995 in the two groups. 20 women in the clock factory and 4 telephonists have made bone cancer.**
  - **1. Specify the type of the epidemiological study**
  - **2. Calculate prevalence, OR and relative risk.**
  - **3. Interpret the results.**

Practical application

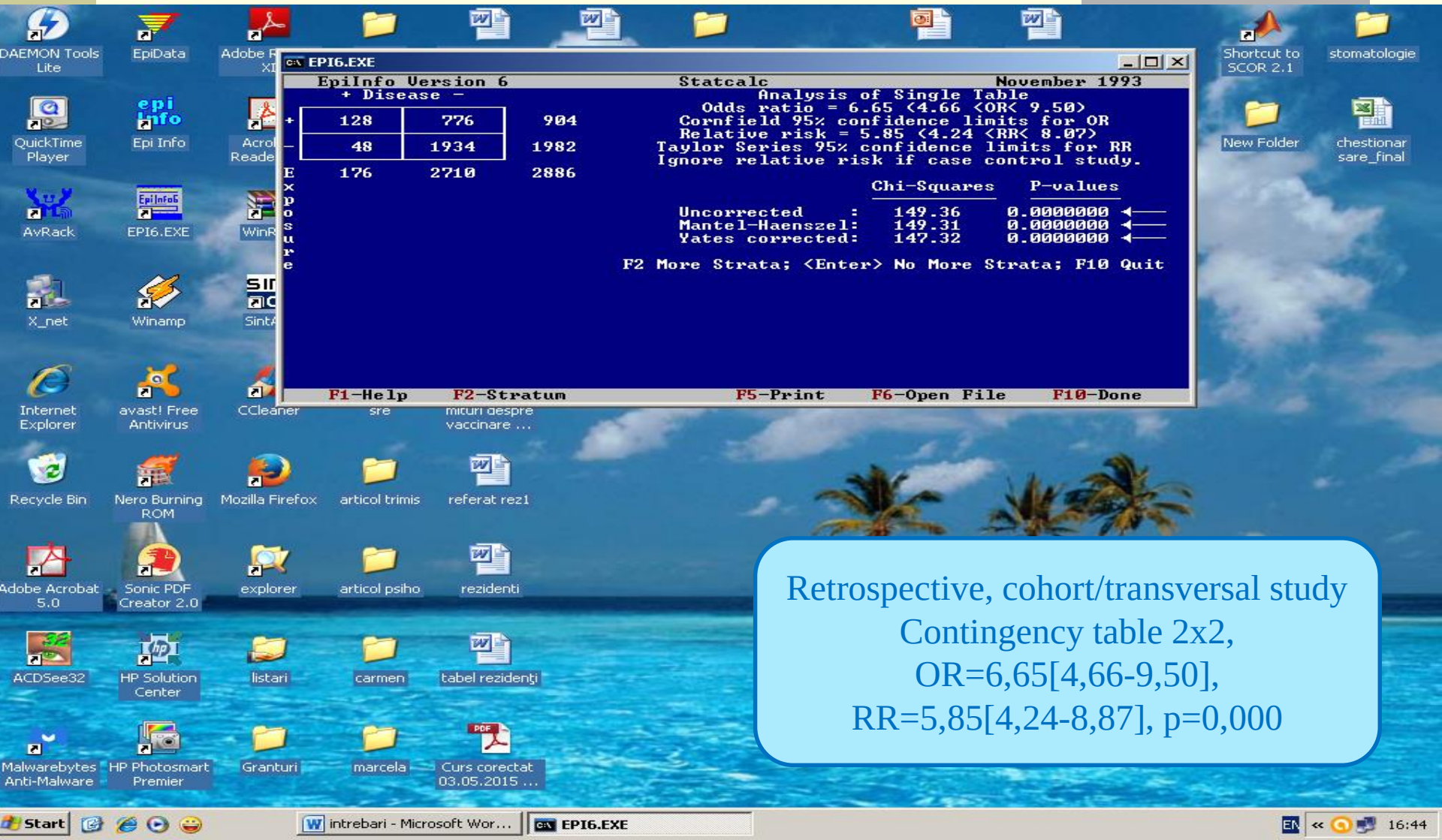


Cohort, prospective study,
Prevalence of 2%/0,4%
OR=5,08 [1,69-20,51],
RR=5,00 [1,72-14,58], p=0,001

Practical application

- **In a survey based on a questionnaire made among high school students in Timiş County in 2005 it was found that 176 pupils out of a total of 2886 had suicide attempts in the last 12 months. 128 students who had suicidal attempts said they felt sad or hopeless every day for two or more consecutive weeks, compared to 776 students who declared the same condition but did not have that attempts.**
  - **Specify the type of the epidemiological study**
  - **For processing the survey data, which type of table is recommended to be used?**
  - **Based on the data obtained from this survey, can you say there is an association between sadness and attempted suicide? In the affirmative, please state how strong this association is.**

Practical application



Retrospective, cohort/transversal study
Contingency table 2x2,
OR=6,65[4,66-9,50],
RR=5,85[4,24-8,87], p=0,000

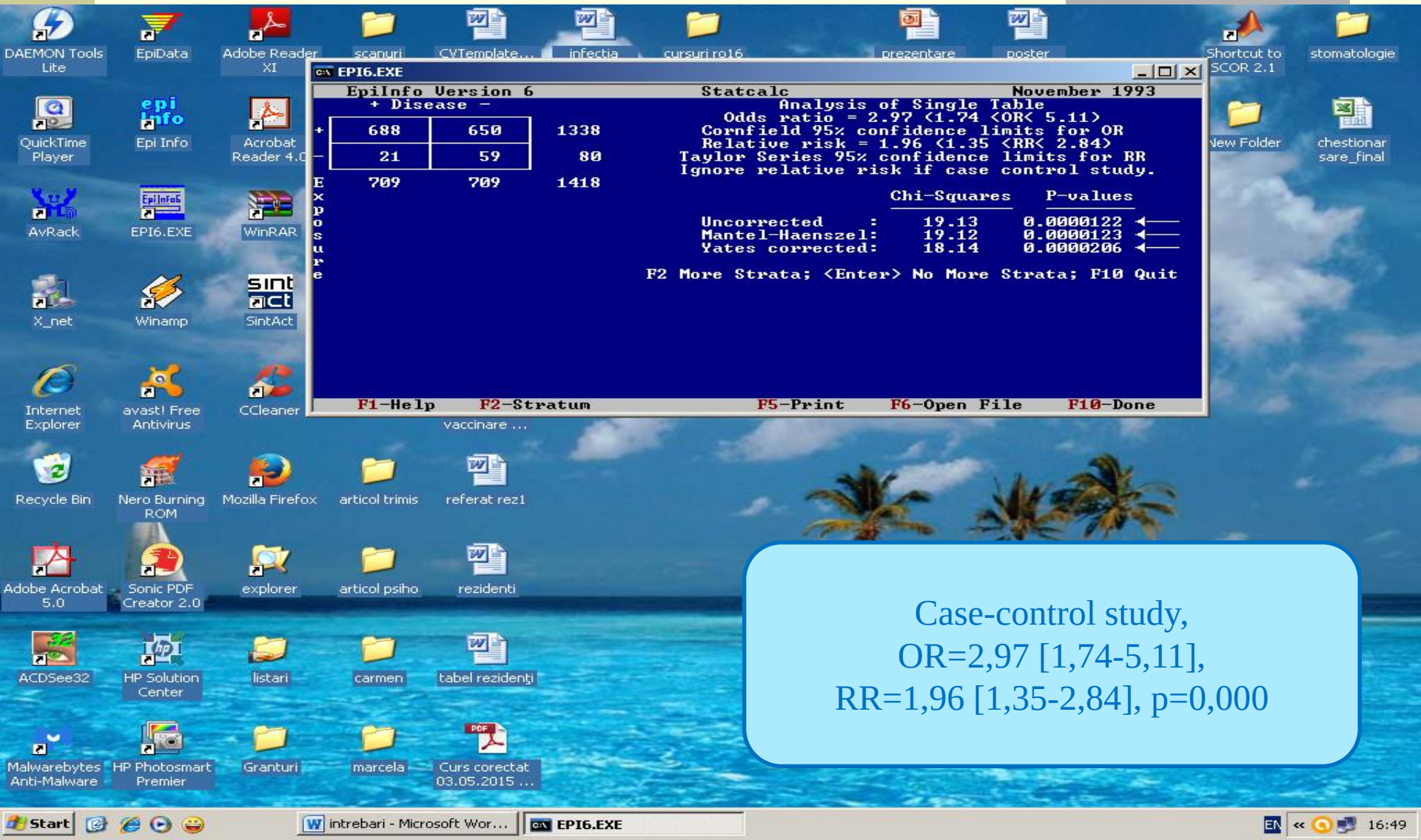
Practical application

In the 1950s, in London, Doll and Hill studied 709 lung cancer patients and 709 non-cancerous witnesses. Four social assistants interviewed hospitalized patients using a set of questionnaires. For each lung cancer patient, the social worker was instructed to interview the first patient on the admission list having the same sex and age group (5 years) and a non-cancer diagnosis as a control patient. For all patients, the discharge diagnosis was checked.

It was found that only 2 of the 649 male lung cancer patients were non-smokers compared to 27 in the witness group. Similarly, only 19 of the 60 female lung cancer patients were non-smokers, compared with 32 in the case of the witnesses. These differences proved to be statistically significant.

- **What type of epidemiological study is this?**
- **Have you concluded, based on these data, that there is an association between smoking and lung cancer?**

Practical application

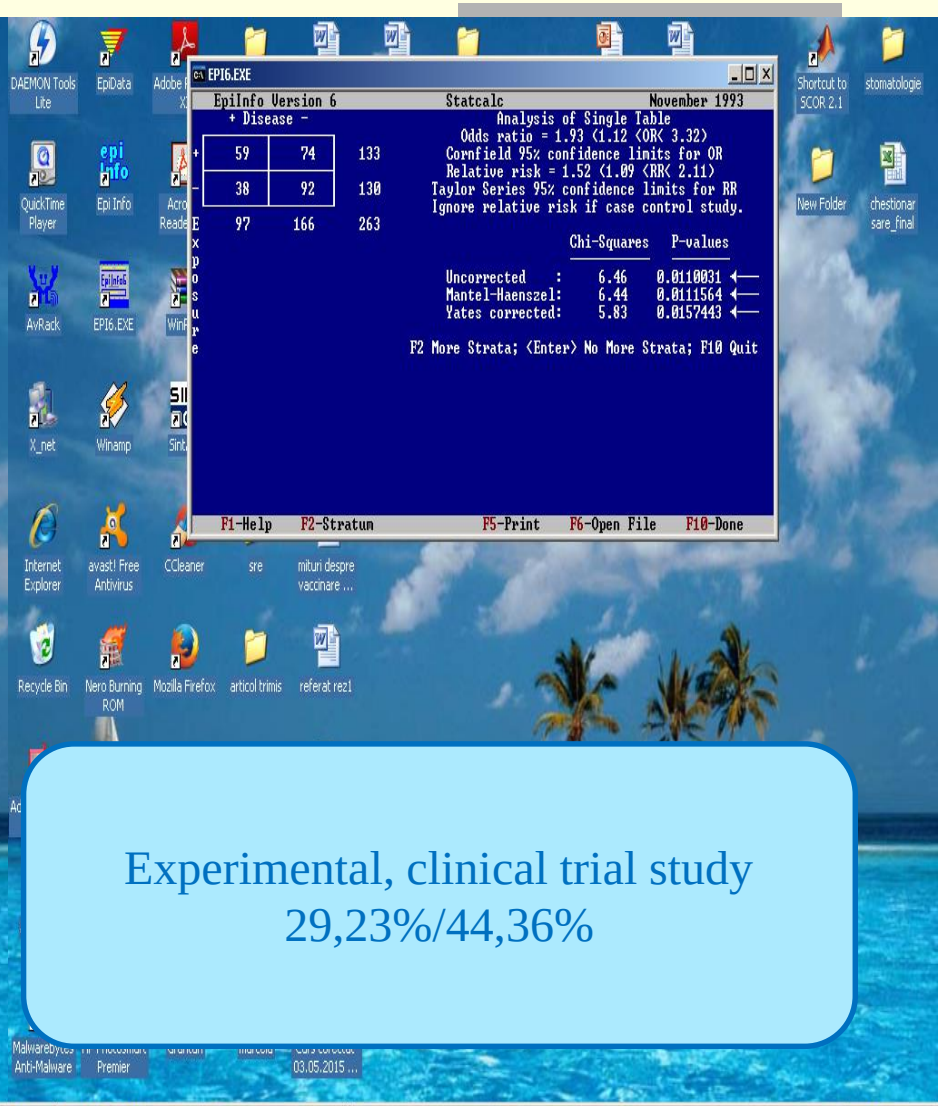


Case-control study,
OR=2,97 [1,74-5,11],
RR=1,96 [1,35-2,84], p=0,000

Practical application

- **The following study was done to assess the effectiveness of targeted therapy on specific targets prior to admission to an intensive care unit. Patients who have arrived at an emergency department with severe sepsis or toxic septic shock received 6-hour treatment to achieve certain goals or standard therapy (as witnesses) before being hospitalized in the intensive care unit . Of the 263 patients enrolled in the study, 130 were randomized to targeted therapy, and 133 were assigned to receive standard therapy. Within 7-72 hours since the admission, 38 in-hospital deaths were recorded in the patients receiving the targeted treatment compared to 59 deaths in those receiving the standard therapy.**
  - **Specify the type of the epidemiological study;**
  - **Calculate death rates in the two groups;**
  - **Calculate OR, relative risk;**
  - **Interpret the results.**

Practical application



Experimental, clinical trial study
29,23%/44,36%

Thank you !