Repetition of Epi I

Gustaf Edgren, PhD

Outline

- Epidemiological basics
- Measures of disease frequency/occurrence
- Measures of association
- Study design (cohort, case-control and intervention studies
- Bias
- Confounding

What do YOU want to repeat?

What is epidemiology?

- The study of the distribution and determinants of disease frequency in human populations
- Key words:
  - Distribution
  - Determinants
  - Frequency
  - Populations

So what are the building blocks of epidemiology?

- Disease frequency:
  - Incidence
  - Prevalence
- Disease determinants = measures of association:
  - Relative risks
  - Absolute risks
- Distribution of disease:
  - Stratification?
  - Risk groups?
- Populations?

Measures of disease occurrence

- Two general types of measures:
  - Prevalence – how many have the condition?
  - Incidence – how many get the condition?

Prevalence

- How many have the disease?
- e.g. 30% of all Americans are obese
- or, 1 in 5 are nearsighted
- Also, recall the difference between point and period prevalence…

Incidence

- How many get the disease?
- Two measures of incidence:
  - Cumulative incidence
    - How many in a predefined population get a certain disease in a specified time-period?
  - Incidence rate
    - How many get a certain disease per unit time?

Examples – Cumulative incidence

- The lifetime risk of breast cancer for US women is 12%
- During the first 5 years of the study period 5% died

Examples - Incidence rate

- The incidence rate of pheocromocytoma in Sweden is approximately 2 per 1,000,000 person years
- The mortality among Swedish AML patients who are blue-collar workers is 389 per 1,000 person years

Pop quiz 1!

- What is the relation between incidence and prevalence?

So, what about disease determinants?

- I suppose a disease determinant is something that determines whether a person gets a disease or not.
- In epidemiology, we often refers to this as a cause, or more frequently a risk factor…
- Usually, we use measure of association for such inference.

Measures of disease association

- There are two main types of disease association:
  - Relative measures:
    - Relative risks, etc.
    - Primarily used when we are trying to make causal inference
  - Absolute measures:
    - Risk differences, etc.
    - Also useful for causal inference, but primarily used when we put on our public health hats

General

- Getting the numbers right is the most important step in understanding the data!
  1. What exposure and outcome are you interested in?
  2. Summarize the data by exposure and outcome
  3. Calculate basic measures of disease occurrence and fitting association measures
  4. Interpret the results!

Epi tool nr 1: The 2-by-2 table

Exposure \ Outcome	+	-	
+	a	b	a + b
-	c	d	c + d

Relative measures

- For count data:
  - Cumulative incidence ratio (CIR) = risk ratios = relative risks (discouraged)
  - Odds ratios (OR) – risk ratio approximation
- For person-time data:
  - Incidence rate ratios (IRR) = rate ratios = relative rate

Absolute measures

- For count data:
  - Cumulative incidence difference (CID) = risk difference = absolute risks (discouraged)
- For person-time data:
  - Incidence rate difference (IRD) = rate difference

Relative and absolute risks

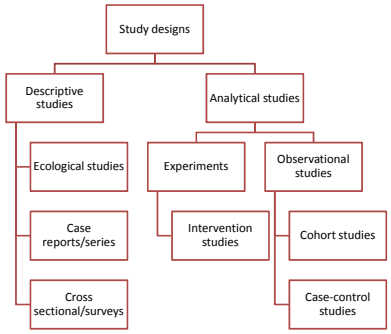
- Relative risks are often used to investigate etiology:
  - Does the exposure cause the disease?
- Absolute risks are often used when we want a public health perspective
  - How important is this exposure for the total disease burden? How much of the disease is caused by the exposure?

"Correlation does not imply causation"

Study design – two principal approaches

Experiments	**Observational studies**
"Change the state of nature and observe the effects"	"Observe nature as it is"
Examples: - Randomized studies - Cross-over studies - Etc.	Examples: - Cohort studies - Case-control studies - Cross sectional studies

Overview – study design



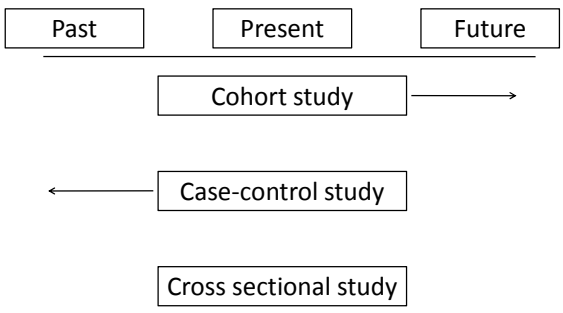
Ecological studies

- The most basic type of descriptive study
  - Based on addressing correlations between variables measures on group level
    - E.g. country level, hospital level, etc.
  - No access to individual data, nor to data on important confounders
  - Often considered the weakest in the "analytical study design hierarchy" (or may be placed among the descriptive study designs)
  - Usually used for hypothesis generation

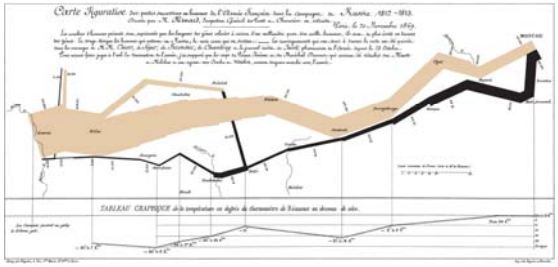
Case reports/series

- Reports of individual or series of cases suggesting a certain causal mechanism
- Naturally only suggestive evidence and is primarily hypothesis generating
- For the excruciatingly rare conditions, sometimes the only available evidence…

Analytical studies, schematic



Cohort studies – mother of all studies



Cohort studies

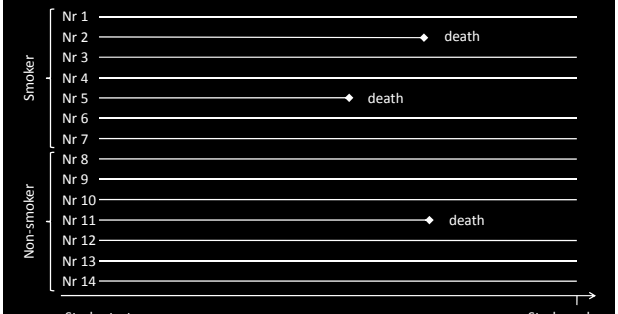
- Cohort studies = prospective/longitudinal studies
- A group of people followed over a period of time for an outcome (in medicine, usually a disease or death)
- The risk of this outcome is correlated to an exposure (or many exposures), which is typically a treatment or a suspected risk factor

Prospective vs. retrospective

- Cohort studies can be conducted using data collective prospectively
- or, with data recorded for other reasons and recycled for the purpose of new research questions
- The distinction is whether the investigator initiates collection of these data (prospective) or uses already collected data (retrospective)
- Thus, the distinction between prospective and retrospective cohort studies is often arbitrary

Cohort studies

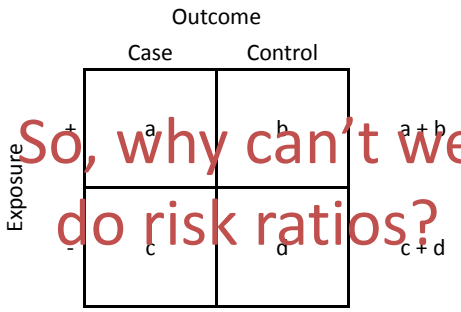
Room for simplification?



Case-control studies

- The case-control study is a clever simplification of the cohort study
- In stead of following individuals over time for disease to occur, we go directly on the diseased and find matching controls
- The goal remains the same, to see how exposures associate with risk…

Association measures in c-c studies?



So, why can't we do risk ratios?

Selection of controls

- The selection of controls is often the most critical element when conducting a case-control study
- Several methods for control selection exist, but the rule of thumb should always be:
  - The controls should be selected from the same **source population** as the cases

Case-control studies – pro and con

+ Possibility to study rare diseases
+ Appropriate for diseases with long induction times
+ Possibility to study any number of exposures
+ Cost effective

– Poor for unusual exposures
– Can be difficult to interpret
– Methodologically more difficult and error prone

The study design hierarchy

- There is a more or less generally accepted evidence hierarchy of the different study designs:
  - RCTs (and variations thereof)
  - Cohort studies
  - Case-control studies
  - Cross sectional studies
  - Ecological studies
  - Case series
  - Case reports

Systematic errors – bias

- In epidemiology, we generally differ between systematic and random error
  - Systematic error is something that occurs in the design of a study and can usually not be mended afterwards
    - (An exception, of course is confounding…)
  - Random error occurs due to natural, random variation and cannot be "controlled." Instead, we estimate its (possible) effects.

Selection bias

- Selection bias occurs when the selection of subjects into a study is **different** depending on the exposure or disease status
  - In a case control study, selection bias occurs when the selection of cases or control is related to the exposure status
  - In a cohort study, selection bias occurs when the selection of exposed and unexposed is related to the (future) disease status

Observation bias

- Also referred to as information bias or misclassification
- Occurs when the characterization of exposure or outcome is imperfect
- Importantly, we differentiate between:
  - Non-differential misclassification – always (?) bias towards the null
  - Differential misclassification – direction of bias is unpredictable

Differential misclassification

- Occurs for the exposure when the degree of misclassification varies according to the outcome status
- Occurs for the outcome when the degree of misclassification varies according to the exposure status

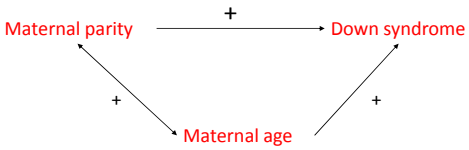
Non-differential misclassification

- Occurs when the misclassification of either the exposure or the outcome is random:
  - i.e. When the degree of misclassification of exposure is not related to the disease status
  - or, when the degree of misclassification of the outcome is not related to the exposure status

Confounding

- The effect of the exposure under study is mixed together with the effect of another variable

Confounding



Confounding factor - definition

- Associated with the disease
- Associated with the exposure
- Not an effect of the exposure

Methods to handle confounding

1. Randomization
2. Restriction
3. Stratification
   - Pooling
   - Standardization
4. Regression analysis
5. Matching (special case)