Common study designs

Zagreb

2019

CHARITÉ BERLIN SCHOOL OF PUBLIC HEALTH

Paper writing: general aspects

- In science, you are what you write
- Good writing cannot overcome bad science
  - But a badly written paper will not get as much attention
- Use definite, concrete, and specific language
- Think how a scientist (ie, you) would read a paper
- Write for the specific readership of the journal you are targeting (ie, general audience, specialty audience)

Paper writing: general aspects

- **STROBE**: Strengthening the reporting of observational studies in epidemiology
- **CONSORT**: Consolidated Standards of Reporting Trials
- **QUOROM**: Quality of Reporting of Meta-analyses
- **MOOSE**: Meta-analysis Of Observational Studies in Epidemiology

Important aspect of reporting is the study design

Paper writing: general aspects, title

- The title of a paper should describe in a few words the content of the paper
- Do not use the conclusion of the paper as the title
- State the main design
  - Migraine and risk of stroke: a case-control study

Paper writing: general aspects, methods

- The Method section should give all aspects of what you did and how you did it
- Use section headers: study population, headache ascertainment, statistical analysis, etc.
- Start writing the Methods section as soon as it is mature

Paper writing: general aspects, methods

- Do not refer to other papers; the paper must stand by itself so readers (and reviewers) do not need to get other papers to understand the methods
- If you use equations, double check, and check again
- After reading the Method section, readers should be able to do the study if they have the data

Paper writing: general aspects, methods

- For studies involving humans, describe how participants were selected and enrolled, and the sites or setting from which they were recruited
- Describe study procedures including any details of interventions (if applicable), measurement and classification of main exposure (if applicable) and outcomes, and other data collection techniques
- Consider the use of a figure to show study processes
- Report how many individuals were eligible, how many declined to participate and how many were lost to follow-up

Common research- study designs

Content

- Introduction to main concepts and strengths/weaknesses in
  - Ecologic
  - Cross-sectional
  - Case-control
  - Cohort
- “Retrospective” vs. “prospective”

Observational studies

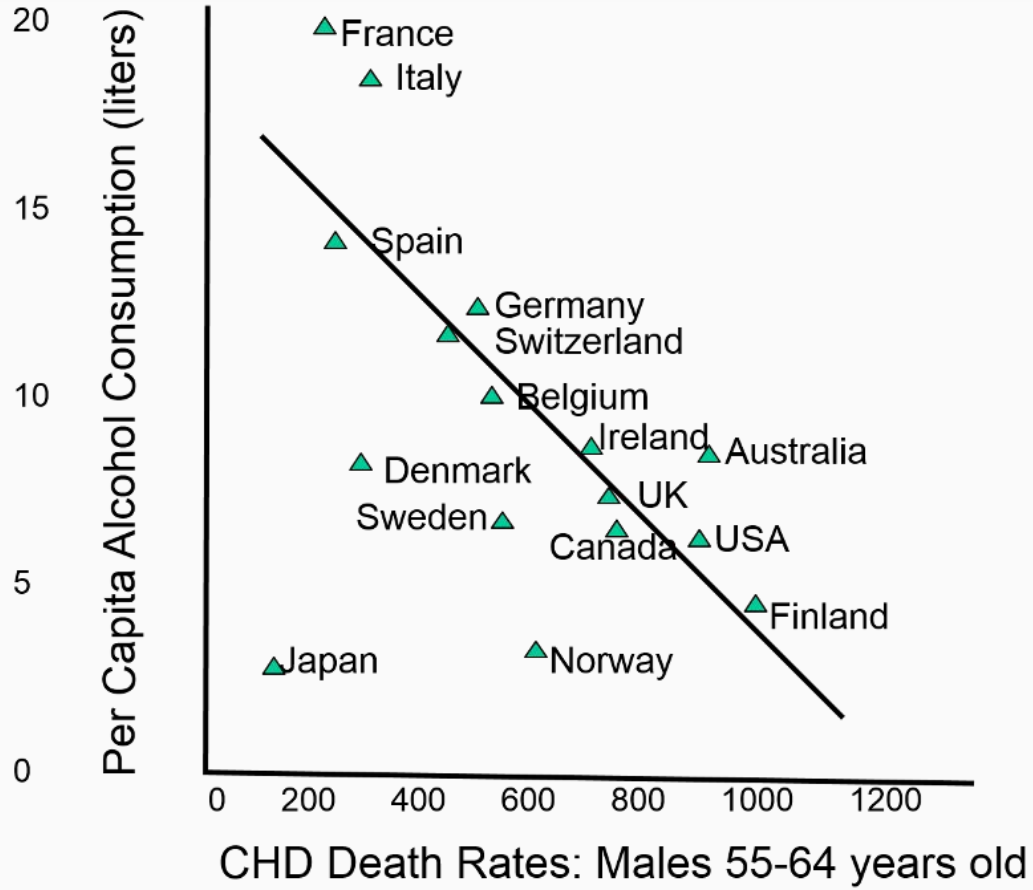
- Descriptive
  - Who? What? Where? When?
  - Correlation or ecologic studies
  - Cross-sectional
- Analytic
  - Why?
  - Case-control
  - Cohort (special form, not observational: randomized controlled trial)
  - (Cross-sectional)

Ecological studies

1. Ecological (correlation) studies

- Descriptive
- Measures that represent characteristics of entire population
- Use aggregate data (e.g. country level)
- Used to describe disease and create hypotheses about possible causal associations
- Measure of interest: correlation between exposure and disease
  - Among different groups
  - Correlation coefficient, r
    - Range -1 to +1
    - 0 = no correlation

“When a correlational study compared per capita alcohol consumption to death rates from coronary heart disease in different countries, it appeared that there was a fairly striking negative correlation.”



example from:
http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_DescriptiveEpi/EP713_DescriptiveEpi7.html

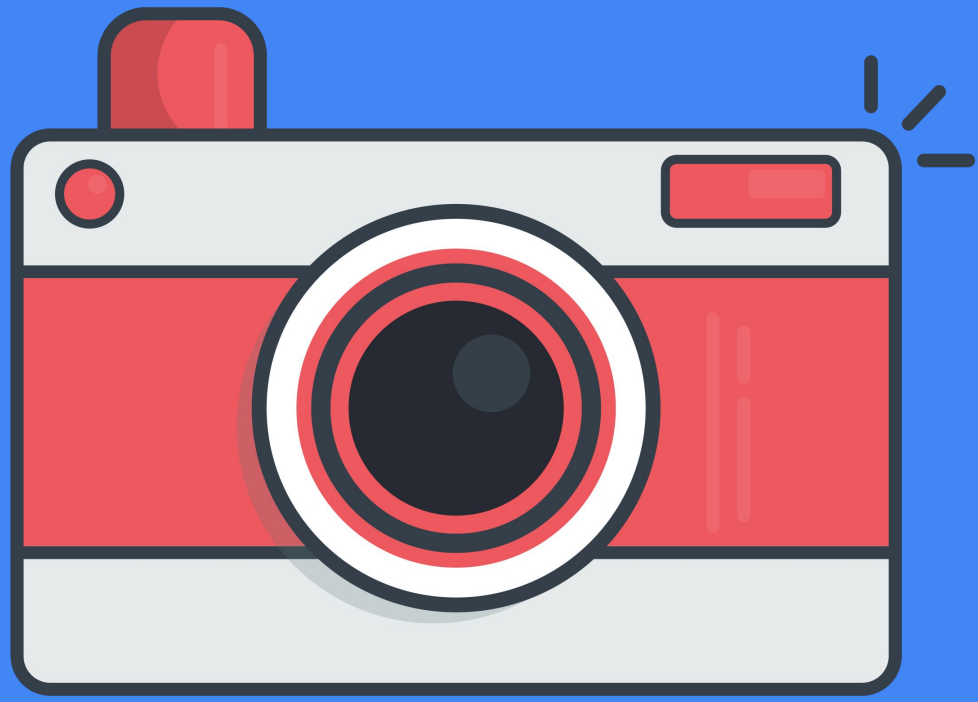
Words of caution…

- Associations on population level may not reflect associations on individual level
  - If we do this = “ecological fallacy” = bias!
  - We can’t directly link the exposure to the disease
- Exposure in correlational studies is the **average** exposure for an entire population or group.

Words of caution…

- We cannot take confounding into account
  - Correlation might mislead us
  - e.g. There may be a number of other differences between the populations that are associated with the exposure
  - Lack of correlation also doesn't imply no association
  - complicated relationships masked in this study design

Cross-sectional studies



2. Cross-sectional studies

- Participants included based on availability at “a point in time” in region, etc. (= ’snapshot’)
- Does not mean that this is done in one day
- No follow-up for the development of the disease
- All information collected at one time point
  - But questions can be asked about the past
- Can be analytic if a clear *a priori* defined cause and effect is studied
  - e.g. effect of a genetic marker on disease status

2. Cross-sectional studies

- Unless exposure develops (biologically) clearly before the disease of interest, strictly **no** inference on temporal sequence possible
  - Example: association between headache and depression
    What is first?
  - Exceptions: genetic markers, conditions develop in childhood or early adults for studies among the elderly

2. Cross-sectional studies

- Common mistake: “in cross-sectional studies, no (causal) effect can be studied just because of the design”
- If you can draw a DAG, you can evaluate an effect!
- The design limits conclusions on the temporal sequence for some study questions

Analytical studies

- Designed to improve on the limitations of the descriptive study designs:
  - among **individuals** (issue with correlational studies),
  - appropriate comparison group (issue with case series)
  - appropriate time sequence (issue with cross-sectional),
  - adequate control of confounding (issue with all studies).

2 types:

- Observational studies
  - (exposures are self-selected or due to environment; investigator passive observer)
  - Case-control and cohort

- Intervention studies (exposures are allocated by investigators)
  - e.g. randomized clinical trials

Case-control studies

3. Case-control studies

- Initial selection of participants based on **disease status**
- Evaluation of **exposure history**

Case-Control Study: Selection into study on basis of disease status

EXPOSURE

DISEASE

PRESENT

ABSENT

Basis on which groups are selected at beginning of study

INVESTIGATOR

3. Case-control studies

- Efficient in design with respect to time and money, as disease already occurred
- Efficient way to deal with long latent period
- Exposure for cancer often cause damage only after several years
- Ideal study design when the outcome is **rare**
- Allows for evaluation of multiple exposures for a single outcome

3. Case-control studies - limitations

- Appropriate data on exposure may be difficult to obtain
- Presence of disease may influence ascertainment of exposure (differential bias, such as recall bias)
- Only for one disease / outcome
- Cannot estimate disease rates (relative odds)
  - Exception: sample fraction of cases and controls known from source population
- **Main issue:** finding appropriate controls
- Difficulty in knowing appropriate time window for assessing exposure and getting accurate past exposure information

3. Case-control studies - control selection

- Purpose of controls is to get **an estimate of the frequency of the exposure(s) in the source population**
- Ideally, the controls are a direct random sample of the source population from which the cases originated
- Controls must be sampled **independently** of exposure
- Can also be done within a cohort study

3. Case-control studies - control selection

- Hospitalized patients
  Advantages:
  - Convenient
  - Inexpensive
  - Cases and control likely similar in accuracy of exposure recall
  - Generally high level of participation

- Can you think of any possible problems?

3. Case-control studies - control selection

- Hospitalized patients - problems?
  - Disease for which controls are hospitalized may be associated with exposure under study
  - Example: Outcome myocardial infarction, controls selected from hospitalized patients for asthma
  - Issue: smoking risk factor for both conditions

3. Case-control studies - control selection

Source (general) population

- Advantages
  - Generally ensures comparability
  - Disadvantages
  - Often difficult to enumerate all members of population as basis for selecting individuals
  - Difficult to gain cooperation for participation
  - Relatively expensive
  - May not recall exposures with same degree of accuracy as cases

3. Case-control studies - control selection

Friends (or family)

- Socialization may be related to some exposures e.g. smoking, alcohol, social isolation, poverty, physical activity, pet ownership …

- The case identifies the control
  - May elect to choose control based on exposure habits:
    Because they are at low risk…
    Because they are at high risk…

3. Case-control studies- design

- A case is matched to a control on disease status (always)
- Sometimes also on age, gender, other factors to reduce confounding by these factors (if also adjusted in the analysis)
- One can match as many controls to the case but due to power considerations
  - Studies have shown: more than 4 is not useful
  - Also: more matched controls= question of feasibility and increased cost
- In general: the matching factor needs to be taken into consideration when **analyzing** case-control studies (conditional logistic regression, 'matching'- more on this later!)

Cohort studies

4. Cohort studies- cohort definition

- Latin: „cohors“ = protected room
- Military: Troupes of the early Roman army

- Epidemiology / Sociology: group of people sharing a defining characteristic / common event / exposure status

4. Cohort studies

- Initial selection of participants based on exposure status

Examples:

  - Occupational groups (nurses, MDs, miners, etc.)
  - Special patient groups
  - Military personnel
  - Geographically defined groups
  - School populations
  - Special groups

Cohort Study: Selection into study on basis of exposure status (observational cohort study)

EXPOSURE

DISEASE

PRESENT

ABSENT

Basis on which groups are selected at beginning of study

INVESTIGATOR

4. Cohort studies- selection of unexposed

Many possibilities:

Usually…

- Internal comparison (unexposed members of the same cohort)

Also:

- Comparison cohort (another cohort from a similar population, that is thought to be unexposed)
- General population data (pre-existing data from the general population, such as National registries)
- Multiple comparison groups

4. Cohort studies- information sources

- Records collected independently of study: occupational, medical, pharmaceutical, education
- Information obtained by research staff: medical exams, environmental or workplace measurements
- Information reported by study subjects (questionnaires, interviews)

4. Cohort studies- warnings

- Caveat: lost to follow-up
  - Differential censoring
- May require complex modeling, in particular if information is collected at several time points during follow-up
- Time-varying information of exposure and covariates in regression models
- Correct assignment of person-time may be challenging
- Internal vs. external validity (**generalizability**)

4. Cohort studies - examples

- Framingham Heart Study (US)
- Nurses‘ Health Study (US)
- Nationale Kohorte (German National Cohort- NaKo)
- Cardiovascular health study (US)
- Berlin Initiative Study (Germany, IPH)

Comparison:

Case-control

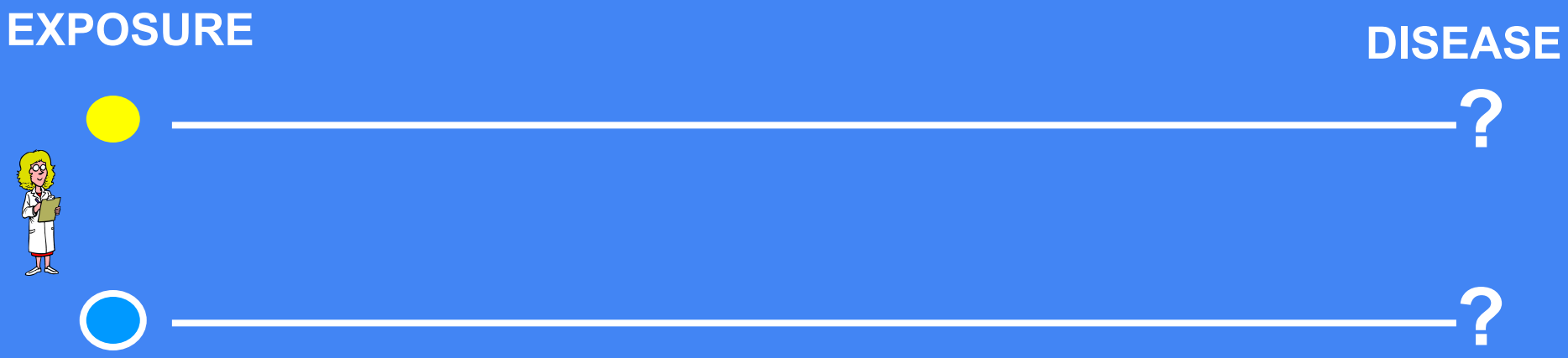
- Selection of participants based on disease status
- Sampling from source population
- Generally less expensive
- Convenient for studying many exposures (but only one disease)
- Can be prospective or retrospective

Cohort

- Selection of participants based on exposure status
- Complete source population as denominator
- Generally more expensive
- Convenient for studying many diseases (and exposures
- Can be prospective or retrospective

Briefly contrast with intervention studies

Intervention Study: Type of prospective cohort study in which exposure is allocated by investigator



PRESENT

ABSENT

Exposure is allocated to participants at beginning of study. Not self-selected; not observational study.

INVESTIGATOR at beginning of study

Intervention study- example 1

- QUESTION: Does immediate treatment after HIV diagnosis (rather than waiting until immune system deteriorates) result in less transmission to sexual partners?
- POPULATION: 1750 couples (one HIV infected, other not) in 14 cities on 4 continents. Allocate to:
- INTERVENTION GROUP: Infected partner put on antiretrovirals as soon as test positive.
- COMPARISON GROUP: Usual care: infected partner start treatment when CD4 count drops below 250.
- OUTCOME: Infection rate of partner during trial with HIV strain genetically proven to come from partner.

Intervention study- example 2

- QUESTION: Does intensive risk factor reduction decrease risk of coronary heart disease (CHD)?
- POPULATION: Multiple Risk Factor Intervention Trial (MRFIT). 12,866 men aged 35 to 57; upper 15% of CHD risk based on smoking, cholesterol, and blood pressure. Allocated to:
- INTERVENTION GROUP: Special intervention program: stop smoking, receive antihypertension medication, and lower cholesterol levels through weight loss or dietary changes.
- COMPARISON GROUP: Usual care.
- OUTCOME: Deaths from CHD.

Prospective vs. retrospective terminology

Retrospective vs. prospective- what's in a name?

- At the time point of study initiation (can be in the past)
- **Retrospective** studies
  - Word means: looking back on or dealing with past events or situations
  - Information on exposure is recorded **after** the outcome has occurred
- **Prospective** studies
  - Word means: expected or expecting to be the specified thing in the future
  - Information on exposure is recorded **before** the outcome occurs
- True for any study design
  - There are retrospective and prospective cohort studies
  - There are retrospective and prospective case-control studies (nested))

Retrospective vs. prospective

- Confusion arises if prospective is interpreted based on the actual timing when a study is conducted (i.e., into the future)
  - Example: Nurses Health Study started in 1972, recorded information on exposure in 1972 and followed people up over time for occurrence of disease
- Retrospective or prospective?

Retrospective vs. prospective

- Confusion arises if prospective is interpreted based on the actual timing when a study is conducted (i.e., into the future)
  - Example: Nurses Health Study started in 1972, recorded information on exposure in 1972 and followed people up over time for occurrence of disease
- Retrospective or prospective?

Prospective, as exposure recorded BEFORE disease occurred!

Retrospective studies

- There is nothing ***wrong*** with retrospective studies
- The main concern is introducing bias when exposure is recorded after the outcome occurred
  - Recall bias, information bias
- Need to understand this: “reality no longer is what it was when it was It cannot be reconstructed by our memories” (Kundera, Ignorance 2003)
  - Without these biases, retrospective as good as prospective

Retrospective vs. prospective

- Only follow-up studies with active treatment assignment are always prospective
- For cohort studies, check when exposure was recorded
- In papers, state: prospective or retrospective **way of data collection or recording**
  - Avoid: “This is a prospective (or retrospective study)” because there are many strong beliefs (misconceptions) about this
  - Example: Migraine and risk of stroke: a cohort study
  - In Methods: information of migraine was collected at baseline and participants were followed for the development of a stroke.

Retrospective vs. prospective- **incorrect** statements

- Retrospective studies are the ‘weaker’ study design
  - Not true if potential biases are ruled out/reduced
- Case-control studies are always retrospective
  - Not true if nested within a prospective cohort study
- Using data from the past means retrospective
  - Not true. Depends if information on exposure is recorded before or after the outcome occurred
- Studies can only be prospective or retrospective
  - Some studies have both a prospective and a retrospective aspect (in subgroups)

Thank you!

Also to Jess Rohmann,
Julie Buring & Pamela Rist