2019 - 2020

Dr. Fazıl Küçük Faculty of Medicine, EMU

Year 1

Biostatistics Course

INSTRUCTOR: Assist. Prof. Dr. İlke Akçay

ilke.akcay@emu.edu.tr

Recall from the previous lecture

- What is statistics and biostatistics?
  - Statistics
    - Collection, organization, summarization, analysis and interpretation of data
    - Population vs. Sample
    - Descriptive Statistics & Inferential Statistics
    - Data
      - Sources of data
- Statistics in medical research

Today's Topics

- Designing Research
- Types of Data

Use of Statistics in Research

PLANNING / literature review, research hypothesis

DESIGN

EXECUTION (data collection)

DATA PROCESSING

DATA ANALYSIS

PRESENTATION

INTERPRETATION

PUBLICATION

The general sequence of steps in a research project

Each step may have substeps

Use of Statistics in Research

PLANNING / literature review, research hypothesis

DESIGN

EXECUTION (data collection)

DATA PROCESSING

DATA ANALYSIS

PRESENTATION

INTERPRETATION

PUBLICATION

The general sequence of steps in a research project

Each step may have substeps

Use of Statistics in Research

PLANNING / literature review, research hypothesis

DESIGN

EXECUTION (data collection)

DATA PROCESSING

DATA ANALYSIS

PRESENTATION

INTERPRETATION

PUBLICATION

The general sequence of steps in a research project

Each step may have substeps

Designing Research

Designing Research

- All medical research is carried out in relation to one or more objectives focusing on the design and plan of the research.
- It is essential to have a good research design.
- Research design is one of the most important aspects of statistical contribution to medicine.

Classification of Scientific Research (Categories of Research Design)

- Based on the collection of data
  - Observational
  - Experimental
- Based on the time relationship
  - Prospective
  - Retrospective
  - Cross-Sectional
- Based on the causal relationship
  - Descriptive
  - Analytical

Classification based on the collection of data: Observational or Experimental

- In an **observational study**, the researcher collects information on the attributes or measurements of interest, but does not influence events.
  - Includes observational surveys and most epidemiological studies.

Classification based on the collection of data: Observational or Experimental

- On the contrary, in an **experimental study**, the researcher deliberately influences events and investigates the effects of the intervention.
  - Experimental studies include clinical trials and many animal and laboratory studies.
  - It is essential to have a **control group** in experimental studies to strengthen the inferences made from the results of the study.
  - Can be performed only if it is feasible and ethically justified to manipulate the postulated cause. (treatments, preventive procedures, health care programs, training)

Classification based on the time relationship

- **Prospective studies**
  - data is collected forwards in time from the start of the study
- **Retrospective studies**
  - data refer to past events and may be acquired from existing sources.

- Experimental studies are prospective but observational studies may be prospective or retrospective.

Classification based on the time relationship

- **Cross-sectional studies** are those in which individuals are **observed only once**.
  - Most surveys are cross-sectional.

Classification based on the causal relationship

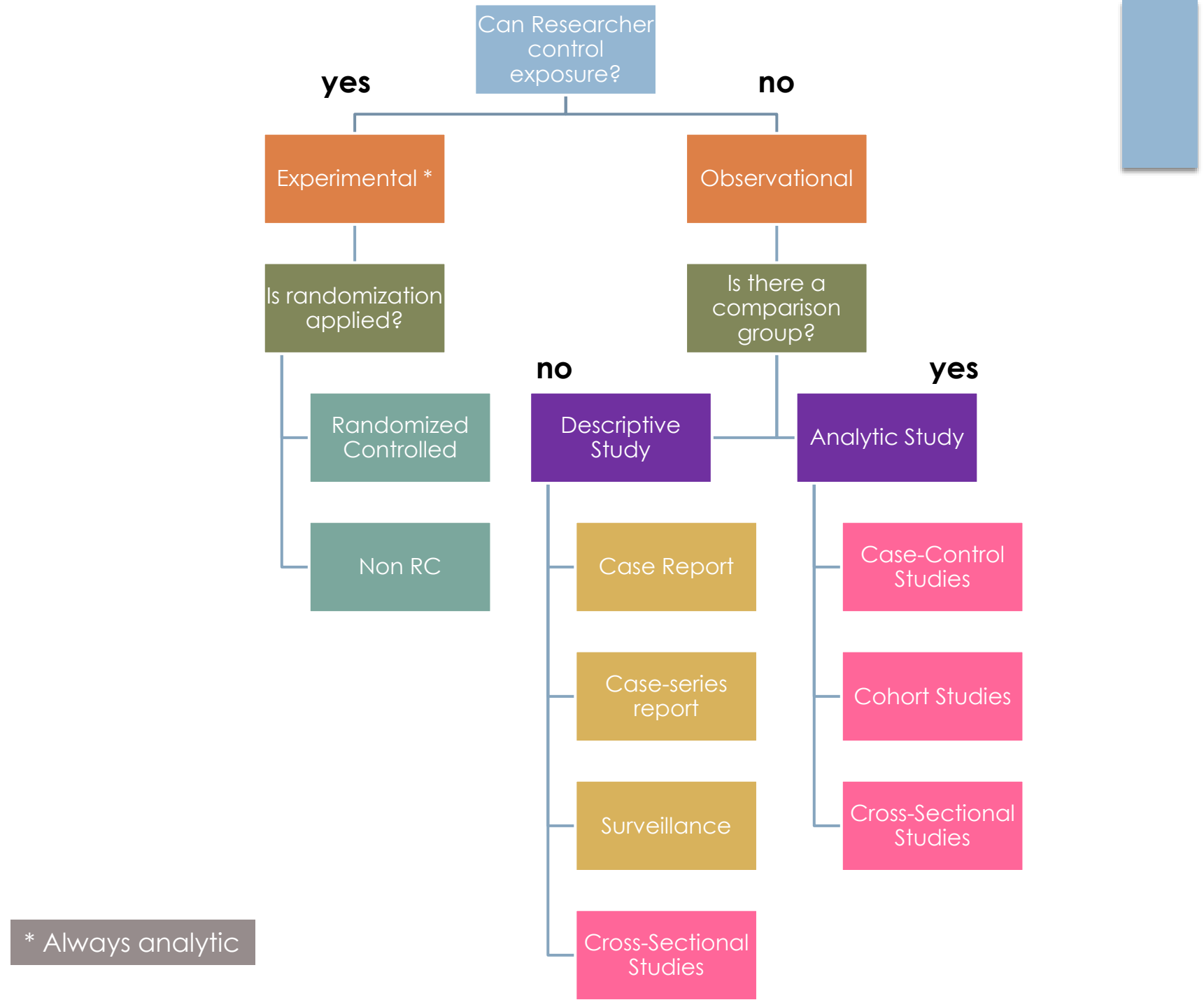
- A **descriptive study** sets out to describe a situation, e.g. distribution of a disease in population relation to sex, age, and other characteristics.

Classification based on the causal relationship

- Descriptive Studies provide
  - insights, data, and information on the course or patterns of diseases, conditions, injuries, disability and death in groups or populations.
  - information about time, place and population characteristics of a health related issue
    - when, where, who
  - clues about cause that can be pursued with more sophisticated research designs (i.e. analytic one)

Classification based on the causal relationship

- An **analytic study** is used to test the exposure-outcome relationship
- There has to be a hypothesis
- There has to be a control or comparison group.


Can Researcher control exposure?
yes
no
Experimental *
Observational
Is randomization applied?
Is there a comparison group?
no
yes
Randomized Controlled
Non RC
Descriptive Study
Analytic Study
Case Report
Case-series report
Surveillance
Cross-Sectional Studies
Case-Control Studies
Cohort Studies
Cross-Sectional Studies
* Always analytic

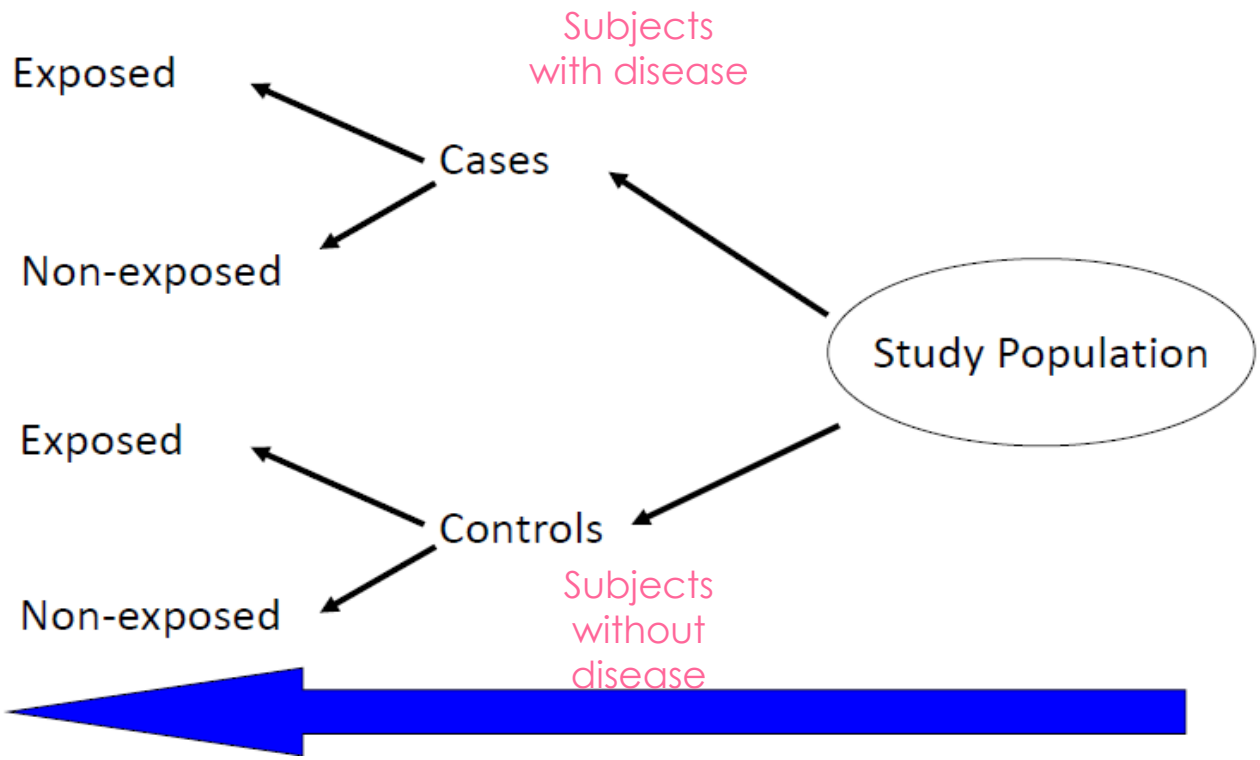
Classification based on the causal relationship

- Analytic Studies
  - Case-control
  - Cohort
  - Cross-sectional
  - Experimental (Intervention) studies
    - Clinical Trials

Case-Control Study

- A **case-control study** starts with the identification of subjects with disease of interest, and a suitable control group of individuals without the disease.
  - The relationship of a risk factor to the disease is examined by comparing the diseased and non-diseased with regard to how frequently the risk factor is present.
  - If the variable under consideration is quantitative, the average levels of the risk factor in the cases and controls are utilized.

Case-control study



Case control studies are retrospective

PASSIVE SMOKING AND LUNG CANCER

PELAYO CORREA LINDA WILLIAMS PICKLE
ELIZABETH FONTHAM YOUPING LIN
WILLIAM HAENSZEL

Department of Pathology, Louisiana State University Medical Center, New Orleans, Louisiana; Environmental Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; Illinois Cancer Council, Chicago, Illinois, USA

Summary Questions about the smoking habits of parents and spouses were asked in a case-control study involving 1338 lung cancer patients and 1393 comparison subjects in Louisiana, USA. Non-smokers married to heavy smokers had an increased risk of lung cancer, and so did subjects whose mothers smoked. There was no association between lung cancer risk and paternal smoking. The association with maternal smoking was found only in smokers and persisted after controlling for variables indicative of active smoking. It is not clear whether the results reflect a biological effect associated with maternal smoking or the inability to control adequately for confounding factors related to active smoking. This preliminary finding deserves further investigation.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Statins and the Risk of Colorectal Cancer

Jenny N. Poynter, M.P.H., Stephen B. Gruber, M.D., Ph.D., M.P.H., Peter D.R. Higgins, M.D., Ph.D., Ronit Almog, M.D., M.P.H., Joseph D. Bonner, M.S., Hedy S. Rennert, M.P.H., Marcelo Low, M.P.H., Joel K. Greenson, M.D., and Gad Rennert, M.D., Ph.D.

ABSTRACT

BACKGROUND

Statins are inhibitors of 3-hydroxy-3-methylglutaryl coenzyme A reductase and effective lipid-lowering agents. Statins inhibit the growth of colon-cancer cell lines, and secondary analyses of some, but not all, clinical trials suggest that they reduce the risk of colorectal cancer.

METHODS

The Molecular Epidemiology of Colorectal Cancer study is a population-based case–control study of patients who received a diagnosis of colorectal cancer in northern Israel between 1998 and 2004 and controls matched according to age, sex, clinic, and ethnic group. We used a structured interview to determine the use of statins in the two groups and verified self-reported statin use by examining prescription records in a subgroup of patients for whom prescription records were available.

RESULTS

In analyses including 1953 patients with colorectal cancer and 2015 controls, the use of statins for at least five years (vs. the nonuse of statins) was associated with a significantly reduced relative risk of colorectal cancer (odds ratio, 0.50; 95 percent confidence interval, 0.40 to 0.63). This association remained significant after adjustment

ORIGINAL ARTICLE

Statins and the Risk of Colorectal Cancer

METHODS

The Molecular Epidemiology of Colorectal Cancer study is a population-based case–control study of patients who received a diagnosis of colorectal cancer in northern Israel between 1998 and 2004 and controls matched according to age, sex, clinic, and ethnic group. We used a structured interview to determine the use of statins in the two groups and verified self-reported statin use by examining prescription records in a subgroup of patients for whom prescription records were available.

RESULTS

In analyses including 1953 patients with colorectal cancer and 2015 controls, the use of statins for at least five years (vs. the nonuse of statins) was associated with a significantly reduced relative risk of colorectal cancer (odds ratio, 0.50; 95 percent confidence interval, 0.40 to 0.63). This association remained significant after adjustment

and verified self-reported statin use by examining prescription records in a subgroup of patients for whom prescription records were available.

RESULTS

In analyses including 1953 patients with colorectal cancer and 2015 controls, the use of statins for at least five years (vs. the nonuse of statins) was associated with a significantly reduced relative risk of colorectal cancer (odds ratio, 0.50; 95 percent confidence interval, 0.40 to 0.63). This association remained significant after adjustment

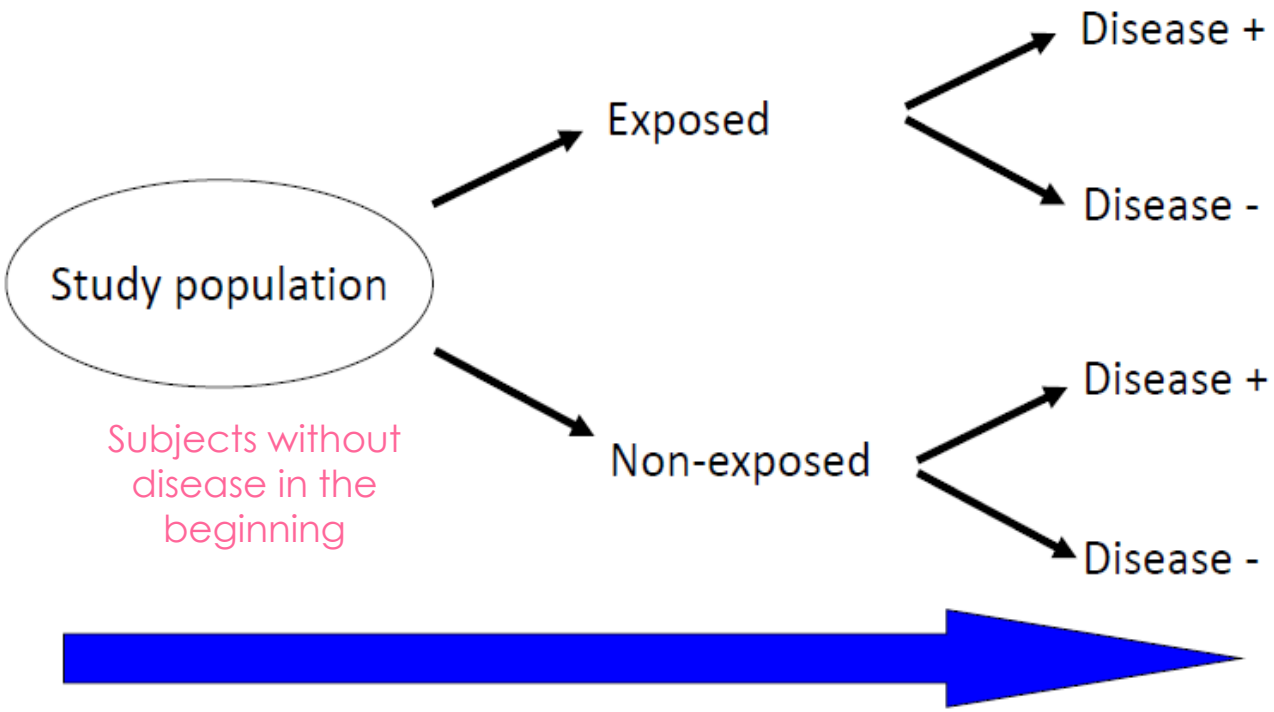
Cohort Study

- A **cohort** is a component of a population identified so that its characteristics (for example, causes of death or numbers contracting a certain disease) can be ascertained as it ages through time.
- The term 'cohort' is often used to describe those born during a particular year but can be extended to describe any designated group of persons who are traced over a period of time.
  - For example, we may refer to a cohort born in 1950, or to a cohort of people who ever worked in a particular factory.

Cohort Study

- A **cohort study**, which may also be referred to as a follow-up, longitudinal or prospective study, is one in which subsets of a defined population can be identified who have been exposed (or will be exposed) to a factor which may influence the probability of occurrence of a given disease or other outcome.
  - A study may follow two groups of subjects, one group exposed to a potential toxic hazard, the other not, to see if the exposure influences, for example, the occurrence of certain types of cancers.
  - Cohort studies are termed observational studies, since they simply *observe* the progress of individuals over time.

Cohort study / Follow-up study



Cohort studies can be prospective or retrospective.

When and why children first start to smoke?

Int J Epidemiol. 1990 Jun;19(2):323-30.
Swan AV, Creeser R, Murray M.

Abstract

Most investigations of smoking in children focus on prevalence in which uptake and maintenance are confounded. This paper reports an analysis of pure incidence data in a cohort of over 6000 Derbyshire schoolchildren followed for ten years investigated using survival data analysis techniques. Over 70% of the cohort tried at least one cigarette before the end of the fifth year of secondary school. Some 40% identified themselves as regular smokers while at school. The risks of taking up regular smoking were higher if, at the age of 11.7-12.7 years, the children had smoking siblings, opposite sex friends, were dismissive of the health hazards and susceptible to peer pressure. More girls than boys in that age range spent time with opposite sex companions and in organized social activities which in turn were significantly associated with the risk of taking up smoking. Thus the earlier physical and emotional development of girls may help explain recent findings that adolescent girls are now more likely to smoke than boys of the same age. The greatest incidence of regular smoking occurred when the average age was increasing from 14.2 to 15.2 years. This has very clear implications for the timing of anti-smoking interventions.

Example from the literature: Cohort study – mortality in slate workers

Campbell et al (2005) studied a cohort of people living in towns in North Wales associated with slate mining. Over a period of 24 years they followed up a group of men who worked or had worked with slate and a control group comprising men who had never been exposed to slate dust. The slate workers and controls were well matched for age and smoking habit. The results are given in Table 12.3.

The risk of death to slate-workers is 379/726 = 0.52 while that for the non-slate workers is 230/529 = 0.43, giving $RR = 0.52/0.43 = 1.21$ with 95% CI 1.07 to 1.36 (see Section 12.11).

In fact the published report took note of the actual survival times and so analysed this study using the survival techniques of Chapter 10. They found a hazard ratio $HR = 1.24$, (95% CI 1.04 to 1.47, $p = 0.015$) and concluded that exposure to slate increases a man's risk of death at any point in time by about 25%.

It is useful to note that in this study the RR and HR are numerically close.

Cross Sectional Studies

- A cross-sectional study
  - describes a group of subjects at one particular point in time.
  - aims to take a ‘snapshot’ of some situation at some particular point or interval in time.
- Non-directional studies
- Data is collected only once from each participant

Cross Sectional Studies

- It may feature the proportion of people with a particular characteristic, which is a prevalence study.
  - It may look at how the prevalence varies by other features such as by age or gender.
- Examples
  - Turkey Demographic and Health Survey
  - Think about diabetes- obesity relationship in a cross-sectional study

Cross Sectional Studies

- Cross-sectional studies are used both descriptively and analytically.
  - Descriptive cross-sectional studies simply characterize the prevalence or incidence of a health outcome (disease, disorder, etc.) in a specified population.
    - For example,
      - the proportion of schoolchildren with iron deficiency
      - the incidence of lung cancer in a year
      - the prevalence of Type-I Diabetes patients in a population

Cross Sectional Studies

- Cross-sectional studies are used both descriptively and analytically.
  - Analytic cross-sectional studies
    - attempts to address more complex situations than descriptive cross-sectional studies
    - investigates potential linkages between variables
    - compares two or more variables among groups of individuals
  - Some examples of analytic cross-sectional studies:
    - Examining the association between mental health outcomes and social determinants in refugees
    - Establishing the impact of age and sex on primary preventive treatment for cardiovascular disease in primary care

Cross Sectional Studies

- Frequently used terms used in cross-sectional studies:
  - Prevalence
  - Incidence
  - Knowledge
  - Awareness
  - Perceptions
  - Attitudes

Categories of Research Design - Summary

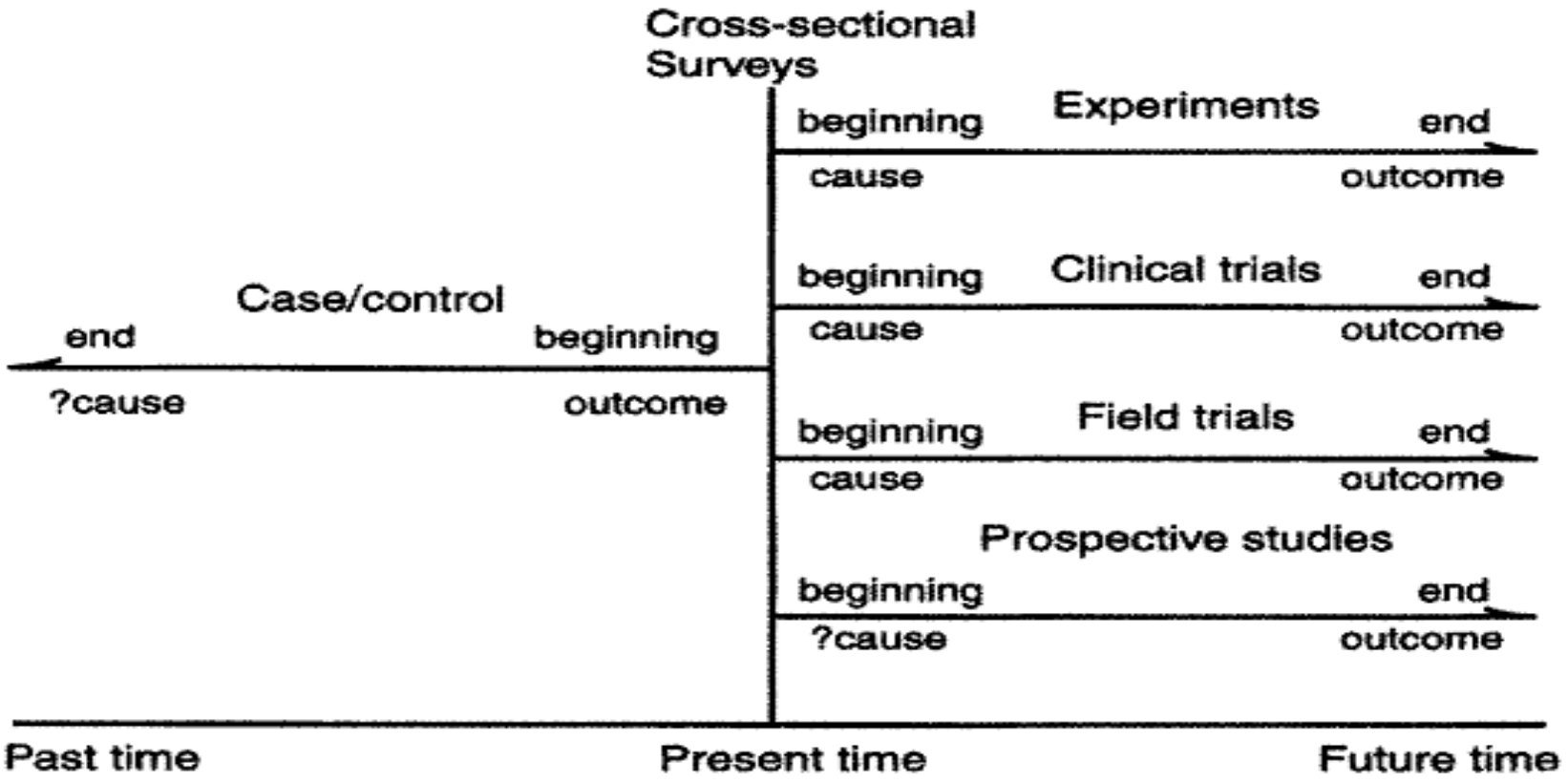


Figure 1.3 Time relationships in biomedical studies.

Types of Data

Types of Data

- Types of data
  - Concept of ‘variable’
    - Qualitative/quantitative variables
    - Discrete/continuous random variables
    - Dependent/Independent variables
    - Measurement Scales
      - Nominal/Ordinal/Interval/Scale

Types of Data / Concept of 'Variable'

- In order to describe types of data, we need to be familiar with the concept of variables.
- The term "variable" is used to describe a quantity that can vary (i.e., take on various values), such as age, height, weight, or sex.
- Variables can be characteristics of a population, such as the age of a randomly selected individual in the U.S. population.
  - Concept of 'variable'
    - Qualitative/quantitative variables
    - Discrete/continuous random variables
    - Dependent/Independent variables

Types of Data / Concept of 'Variable'

- Variable is a characteristic that takes on different values in different persons, places, or things.

Example:

- heart rate,
- the heights of adult males,
- the hair color of preschool children,
- the ages of patients seen in a dental clinic.

- Types of variables:
  - Quantitative
  - Qualitative

Types of Data / Concept of 'Variable'

- Quantitative & Qualitative Variables

Quantitative Variables

It can be measured in the usual sense.

(Numerical)

Example::
- the heights of adult males
- the weights of preschool children
- the ages of patients seen in a dental clinic.

Qualitative Variables

Many characteristics are not capable of being measured. Some of them can be ordered or ranked. (Categorical)

Example:
- classification of people into socio-economic groups
- Hair color
- social classes based on income, education, etc.

Types of Data / Concept of 'Variable'

- Discrete & Continuous Random Variables

Discrete Random Variable

- is characterized by gaps or interruptions in the values that it can assume.

Example:

- The number of daily admissions to a general hospital,
- The number of decayed, missing or filled teeth per child in an elementary school.

Continuous Random Variable

can assume any value within a specified relevant interval of values assumed by the variable.

Example:

-Height,
-weight,
-skull circumference.

No matter how close together the observed heights of two people, we can find another person whose height falls somewhere in between.

Types of Data
Concept of 'Variable'

Dependent	Independent
Explained	Explanatory
Response	Predictor

- Dependent & Independent Variables

Dependent Variable

Represents the output or effect
(the main variable that your research is based on)

Independent Variable

Represent inputs or causes

Example: The effects of *antianginal drugs* on *myocardial oxygen consumption*.

antianginal drugs: one of the independent variables;

myocardial oxygen consumption: dependent variable

Types of Data / Concept of 'Variable'

- Variable types according to Measurement Scales:
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Types of Data / Concept of 'Variable'

- A **nominal variable** consists of named categories, with no implied order among categories.
  - E.g. classifying eye color into 4 categories:
    - Black/Brown/Blue/Green
  - E.g. Male-female; well-sick; >65 yrs - <65 yrs
- An **ordinal variable** consists of ordered categories, where the difference between the categories cannot be considered to be equal.E
  - E.g. In the healthcare field, patients are often rated as;
    - much improved/somewhat improved/same/worse/dead
  - E.g. Socioeconomic status: low, medium, high

Types of Data / Concept of 'Variable'

- An **interval variable** has equal distances between values, but the zero point is arbitrary.
  - E.g. temperature
- A **ratio variable** has equal intervals between values and a meaningful zero point.
  - E.g. Most laboratory test values are ratio variables

Variable Type	Assumptions
Nominal	Named Categories
Ordinal	Same as nominal plus ordered categories
Interval	Same as ordinal plus equal intervals
Ratio	Same as interval plus meaningful zero