What to do today (Jan 17, 2023)?

Part I. Introduction

Part II. Epidemiologic Concepts and Designs

Part II.1 Epidemiologic view of diseases and populations

Part II.2 Measuring disease frequency in population

Part II.3 Overview of Designs for Medical Studies

- **II.3.1 Introduction**
- **II.3.2 Types of Study Design**
- **II.3.3. Related Issues**

Part II.3.1 Overview of Designs for Medical Studies: Introduction

In general, a practical study is as follows

Problem Identification
(Specify question / hypothesis)
$\Downarrow$
Protocol Development
(Select a study design)
$\Downarrow$
Data Collection
$\Downarrow$
Data Analysis
$\Downarrow$
Conclusions

Part II.3.1 Introduction

Often conclusions are made via **descriptive analysis** and/or **statistical analysis**:

- Judgment is required to make inferences from the sampled population to the target population, *when the two populations are different*.
- Random sampling is required to make statistical inferences from the sample to the sampled population, *when the sample is a subset of the sampled population*.

Target Population
$\Downarrow \Uparrow$
Sampled Population
$\Downarrow \Uparrow$
Sample

$\Longrightarrow$ **Study design is very key.**

Part II.3.2 Types of Study Design

Example II.3.1. Data from a 1998 general social survey: a random sample of $n = 1127$ subjects were classified according to presence/absence of two characteristics, yes/no of belief in afterlife and female or male.

	Belief in Afterlife	
Gender	yes	no/undecided
female	509	116
male	398	104

(Agresti, 2007)

$\Longrightarrow$ **study design?**

Part II.3.2. Types of Study Design

Example II.3.2. Data from a Harvard physicians' health study: enrolled subjects were randomized to placebo or aspirin group, and recorded were whether they had any heart attacks during the 5-year study.

	Myocardial Infarction	
Group	yes	no
placebo	189	10,845
aspirin	104	10,933

(Agresti, 2007)

$\Longrightarrow$ **study design?**

Part II.3.2 Types of Study Design

- A **study design** is a plan for selecting study subjects and for obtaining data about them.
- There coulbe be many possible designs; but in practice, a few standard designs account for most epidemiologic research, say, and offer enough flexibility to address a wide range of research questions.
- There are many possible ways to classify study designs, depending on which features are highlighted.

Part II.3.2 Types of Study Design: Observational vs Experimental Studies

- An observational study draws inferences about the possible effect of a "treatment" on subjects, where the assignment of subjects into a "treated" group versus a "control" group is outside the control of the investigator.

  Example II.3.1. the social survey on belief in afterlife

- An experimental study, such as a randomized controlled trial, where each subject is randomly assigned to a "treated" group or a "control" group before the start of the treatment.

  Example II.3.2. the aspirin clinical trial

Part II.3.2. Types of Study Design: Cohort Study

- A **cohort** is a group of people who share a common characteristic or experience within a defined period (e.g., are born, are exposed to a drug or vaccine or pollutant, or undergo a certain medical procedure).
  Example II.3.1 is a cohort study

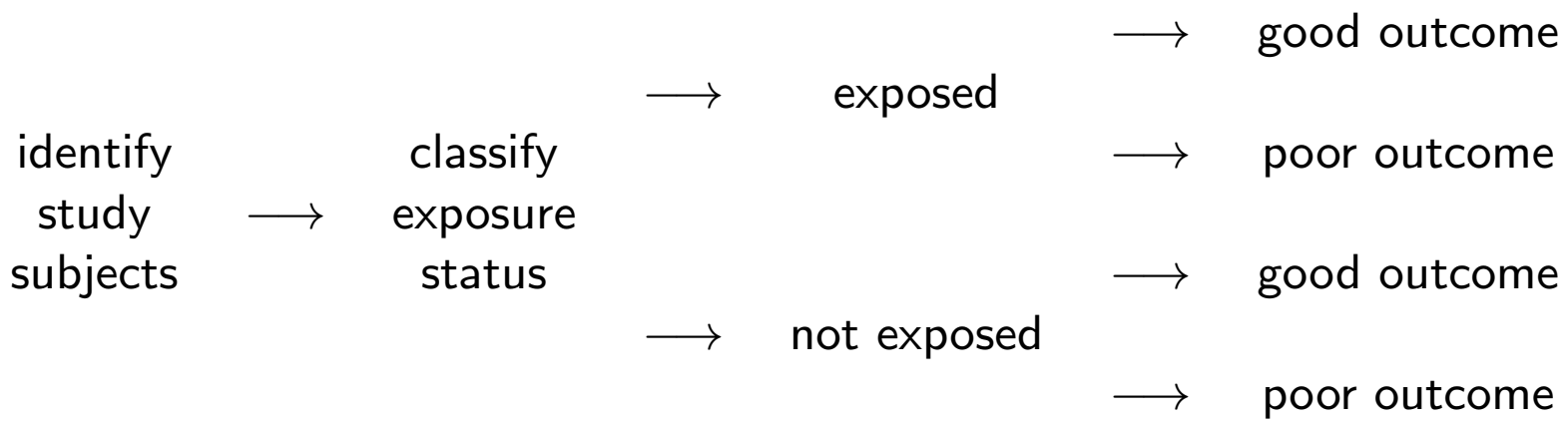


Figure 5-4. Cohort Study (Koepsell and Weiss, 2003)

Part II.3.2. Types of Study Design: Case-Control Study

Often cohort studies are not efficient with rare events/exposures such as cancer/smoking $\Longrightarrow$ another type of study design ...

Example II.3.3 An early study on the association of lung cancer with smoking: a random sample of 709 lung cancer patients, and a random sample of 709 non-lung cancer patients were respectively categorized according to ever smoked or not.

	Have Smoked	
Lung Cancer	yes	not
case	688	21
control	650	59

(Agresti, 2007)

Part II.3.2. Types of Study Design: Case-Control Study

A **case-control** study compares the frequency of past exposure between *cases* who developed the disease and *controls* whom were chosen to reflect the frequency of exposure in the underlying population at risk from which the cases arose.

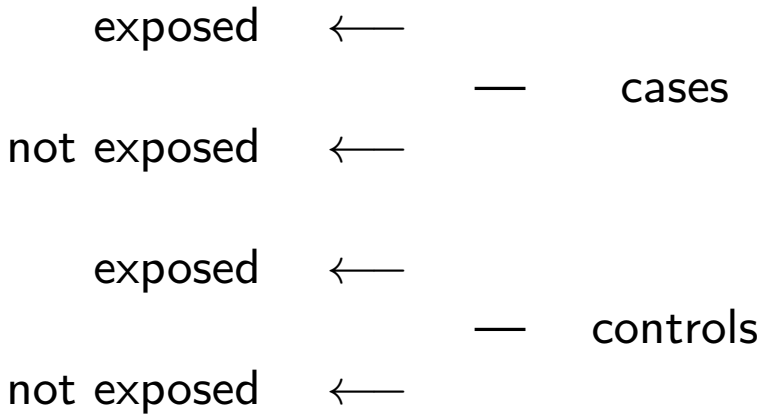


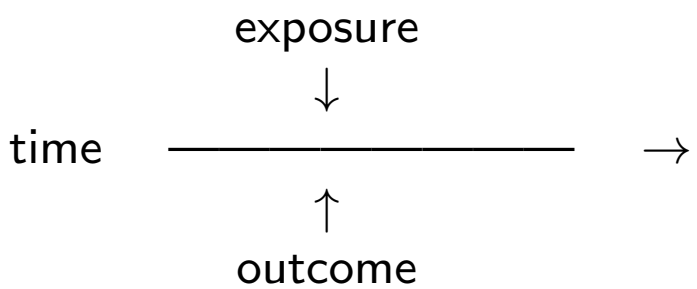
Figure 5-5. Case-Control Study (Koepsell and Weiss, 2003)

What a case-control study can/cannot answer?

Part II.3.2. Types of Study Design: Cross-Sectional Study

A **cross-sectional** study appears to ascertain the exposure and outcome at the same point/period in time.

Figure 5-7. Cross-Sectional Study (Koepsell and Weiss, 2003)



time trend? a cross-sectional clinical trial?

Part II.3.2. Types of Study Design: Longitudinal Study

A **longitudinal** study collects information on exposure/outcome repeatedly over time on study individuals: entitling the capacity of longitudinal studies to separate cohort and time effects

- contrasting case-control studies
- *repeated measures*, a classical version: record of a variable over time
- *time*, a generic term: the metameter for the occasions of observation
  comparing with *spatial data*

Part II.3.2. Types of Study Design: Randomized Trial

A **randomized trial** uses a formal chance mechanism to assign participants either to receive an intervention (or more) of interest or to serve as a control (or more).

- a solid basis for an inference of cause and effect
- *randomization* – control over confounding
- *experimental* vs *control* study arms – intervention vs placebo

more on clinical trial later ... ...
more on study design later ... ...

Part II.3.3. Related Issues – Sources of Data on Disease Occurrence

In the whole population, Common Data Sources for Identification of Disease Cases (Koepsell and Weiss, 2003):

- **Numerator Data**
  - **fatal**: death/fetal death certificates
  - **nonfatal**
    - medical care used: case reports, registry data, clinical records (e.g. hospitalization), administrative data (lab records, drug data)
    - no medical care used: household survey data
- **Denominator Data**: nation/region/division/... census

Part II.3.3. Related Issues – Sources of Data on Disease Occurrence

Uses of multiple data sources ...

- excluding ineligible cases
- estimating the number of missed cases
- reducing misclassification
- expanding opportunities for research through data linkage

Example. the CAYACS program with BC Cancer Agency: A multi-project longitudinal cohort study on *Childhood, Adolescent, and Young Adult Cancer Survivors*.

`https://www.bccrc.ca/dept/ccr/programs/childhood-adolescent-and-young-adult-cancer-survivors-program-cayacs`

- CAYACS before 2018: Mary McBride
- CAYACS II

Part II.3.3. Related Issues – Person, Place, and Time

- **person**: what kinds of people tend to develop the disease, and who tends to be spared? what's unusual about those people? age, gender, race and ethnicity, socioeconomic status, marital status ...

- **place**: where is the disease especially common or rare, and what's different about those places?
  geographic variability

- **time**: how does disease fequency change over time, and what other factors are temporally assoicated with those changes? secular trends (patterns of chance in disease frequency over period of calendar time) – cyclical variation, birth cohort, ...

A descriptive analysis ... ...: 5W Questions

CAYACS Program with BC Cancer Agency

"The CAYACS research program has successfully developed an infrastructure for survivorship research for **s**urvivors of **c**ancer diagnosed in **c**hildhood, **a**dolescence, and **y**oung **a**dulthood, using existing population-based datasets and record linkage methodology. The research program is conducting a series of epidemiologic, clinical, and health services studies relating to survivorship issues in multiple domains (health, education, resource use etc). ..."

Goals ...

- to assess the mortality and morbidity
- to evaluate the demand of health-care resources: hospitalization, physician visit, etc
- to compare with health controls: education, soci-economic status, etc

Changelles ... and statistical research

- recurrent events with non-ignorable duration;
  e.g. Hu et al (2010); Zhao and Hu (2013)

- latent classes;
  – longitudinal vs cross-sectional studies based on physcian claims
  e.g. Wang et al (2014); Wang (SFU PhD thesis, 2016)

- ... ...

Evolving into **Cancer Survivorship Research Program** at BCCA

- correlated multiple event times: semi-competing risks in breast cancer survivors
  e.g. Li (SFU PhD Thesis, 2018), Li et al (2019), Li et al (2021)
- features of administrative data: left-truncation, left censoring?

Remarks: many other studies with health administrative data, resulted from linked different public health databases.

- e.g. Mental Health Related ED Visits (Rosychuk, U of Alberta)
- e.g. Opioid Use Disorders (Nosyk, SFU/St Paul Hospital)

What to study next?

Part I – Introduction

Part II - Epidemiologic Concepts and Designs

- *II.1 Epidemiologic view of diseases and populations*
- *II.2 Measuring disease frequency in population*
- *II.3 Overview of designs for medical studies*

Part III - Clinical Trials

- **III.1 Clinical trial design principles**
- **III.2 Types of clinical trials**
- *III.3 Study monitoring*
- *III.4 A real-life example*

Part IV - Modern Biostatistical Methods