Epidemiology and Study Designs

November 6, 2023

PBHLTH 198, Fall 2023 @ UC Berkeley

Andrew O'Connor

Class Outline

- Announcements: Final Project

- Prelude to Epidemiology

- Prevention, Intervention

- Diseases

- Measures of Disease Occurrence

- Measures of Morbidity, Mortality

- Study Designs

Final Project: Heart Attack!

Logistics

- Final Project (25%); Final Project Reflection (20%) = **<u>45% of your grade</u>**

- Partner project

- Graded 80% on completion, 20% on accuracy

- 22 questions total; 2 are optional

- Total 100 points

  - 10 points extra credit (Max score: 110/100)

- Due **Monday November 27th** by 11:59PM

  - Sorry… will need to reschedule class on the last week (I will be flying back from Taiwan!)

Final Project: Heart Attack!

Project Structure

Section 1: Epidemiology in Practice will introduce you to a practical application of epidemiological research, providing insights into different types of studies, research methodologies, causal inference techniques and the significance of epidemiology in public health.

Section 2: EDA (Exploratory Data Analysis) of CVD (Cardiovascular Disorder) takes you into the realm of data analysis, where you will apply your epidemiological knowledge to real-world health datasets. You'll gain practical experience in understanding and interpreting health data.

Section 3: Predicting Heart Attacks leverages the predictive modeling and machine learning to build a binary classifier, enabling you to predict heart attacks. You will bridge the gap between epidemiological insights and data-driven decision-making in healthcare.

Section 4 (Optional, Extra Credit): Life's Essential 8 gives you the opportunity to compare/contrast the lived realities of Americans in regards to suggested behaviors that promote "good" health. Public health officials may recommend and promote certain behaviors but socially and economically, are these feasible? This section lets you explore this.

Any questions?

Epidemiology Intro

What is epidemiology?

Types of Prevention

The Epidemiologic Approach

Epidemiology

- Epidemiology is the science of public health

- Tells us about disease or health outcomes

- Provides scientific base for testing interventions to prevent and treat disease

– Who (is affected)?
– What (is condition)?
– Where (does it occur)?
– When (does it occur)?
– Why (does this occur)?
– How (is it transmitted, does it develop, and how can it be prevented)?
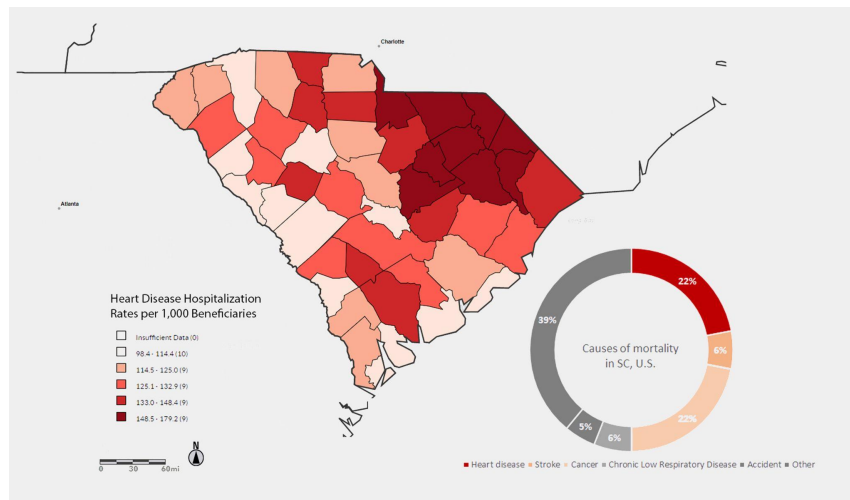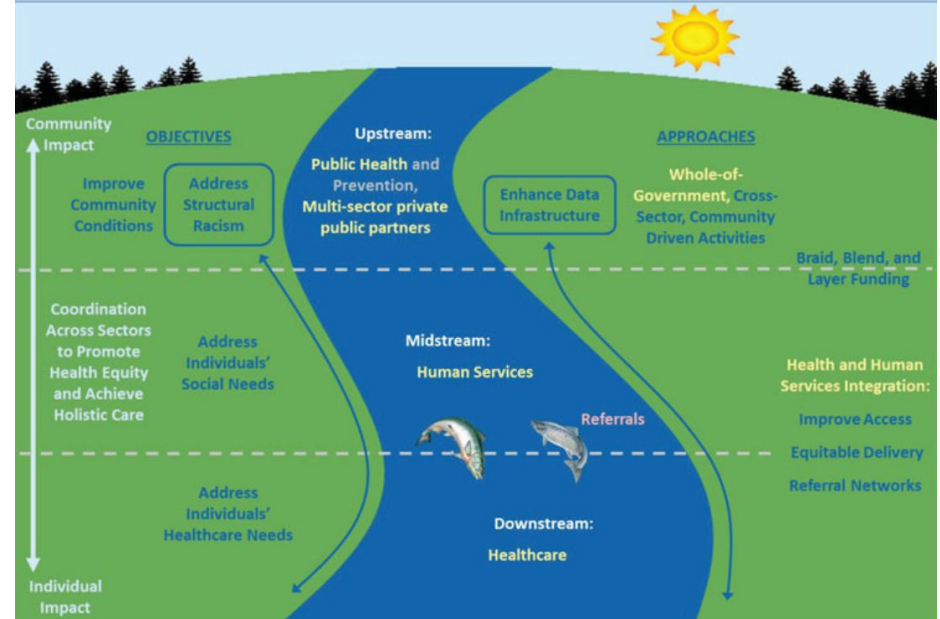
[Lecture 1] Public Health

Public health is "the science and art of **preventing** disease, **prolonging** life, and **promoting** health through the organized efforts and informed choices of society, organizations, public and private communities, and individuals." - CDC

Key Words:
- **Upstream**
- **Disease Prevention**
- **Primary Prevention**

Public Health: How can we prevent people from going to the doctor's office? How can we encourage people to live healthy and fulfilling lives? How can we make this equitable for all?

Types of Interventions

- **Primary**: Prevent the initial development of disease.
  - Educate individuals and communities to avoid exposure; immunization
- **Secondary**: Early detection of existing disease. Screen or survey populations (ideally "high risk" populations) to find early disease and treat before it becomes established; use surveillance systems to identify high risk groups
- **Tertiary**: Reduce impact of a disease. Treat established disease to avoid progression and complications of disease
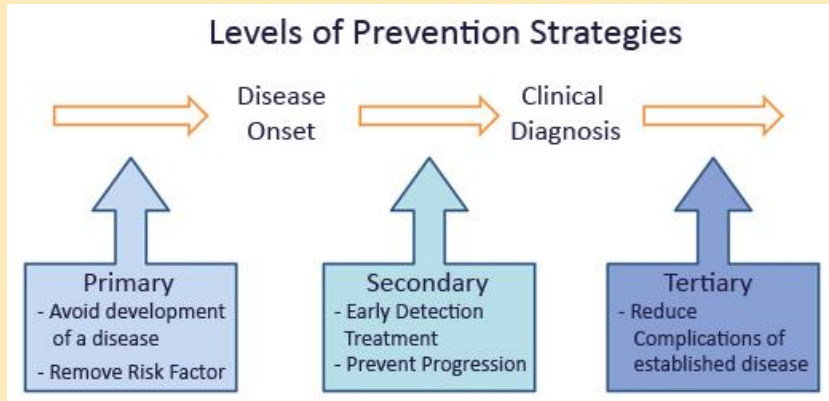
Preventing Tuberculosis
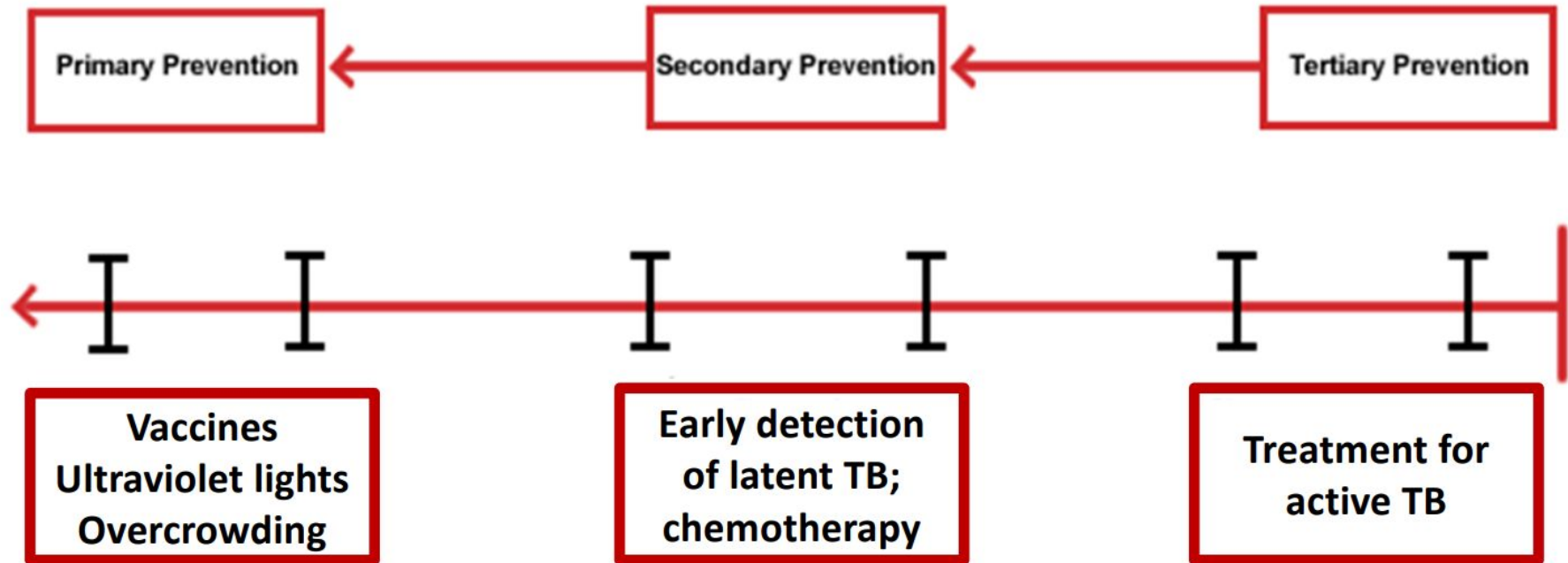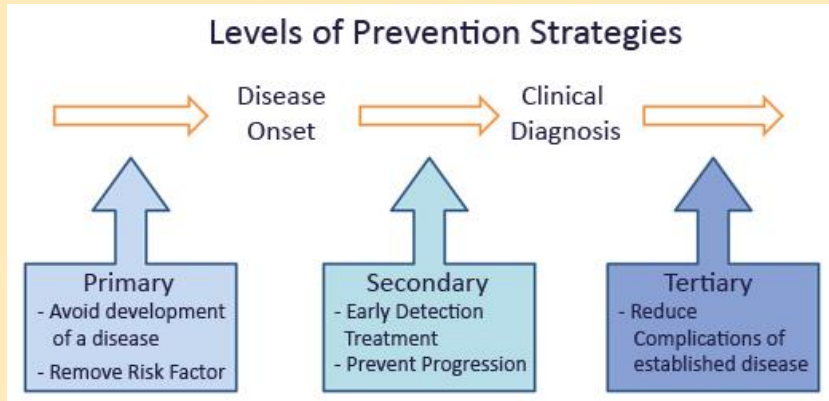
Levels of Prevention Strategies

Exercise: Interventions

Determine the best type of intervention.

Scenario 1: A new strain of influenza is spreading rapidly within a community.

Scenario 2: A clinic offers counseling and support services to individuals who have already developed substance abuse issues and are seeking help to overcome their addiction.

Scenario 3: A medical center offers regular mammography screenings for women over the age of 40 to detect breast cancer in its early stages

Levels of Prevention Strategies

Exercise: Interventions

Determine the best type of intervention.

Scenario 1: A new strain of influenza is spreading rapidly within a community.
Primary

Scenario 2: A clinic offers counseling and support services to individuals who have already developed substance abuse issues and are seeking help to overcome their addiction.

Tertiary

Scenario 3: A medical center offers regular mammography screenings for women over the age of 40 to detect breast cancer in its early stages

Secondary

The Epidemiologic Approach

1. To describe the health status of populations

    - enumerate the cases of disease

    - obtain frequencies of disease within subgroups

    - discover important trends in the occurrence of disease

2. To explain the etiology and natural history of disease

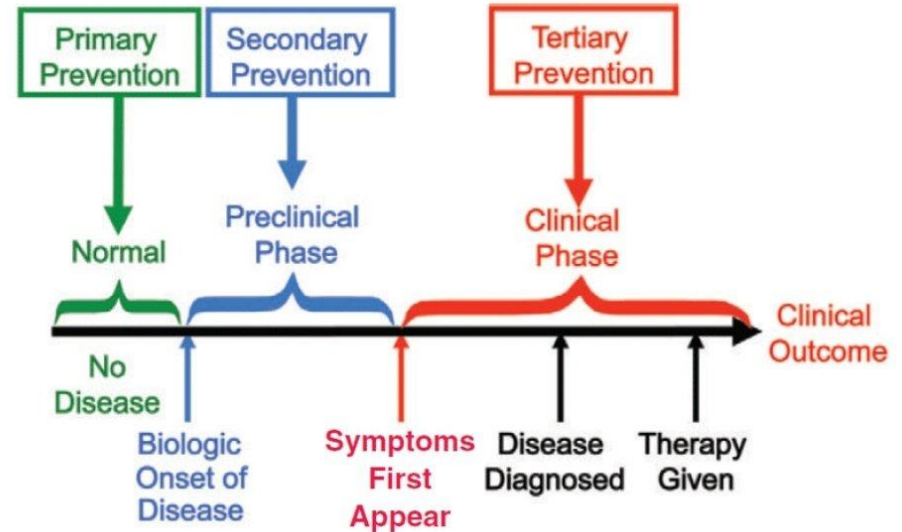   - Discover causal factors; determine modes of transmission

etiology = cause

Diseases

Natural History of Disease

The Natural History of Disease

- Diseases progress in humans in stages
- The different stages of disease over time is called the **natural history** of a disease

Natural History of (untreated) Syphilis

Definition

- **Syphilis**: sexually transmitted disease (STD) or congenital transmission (mother to child) caused by Treponema palladium spirochete bacterium that can cause long term complications if not treated

Symptoms

- painless chancres 10-90 days after potential exposure
- rash all over body and or flu like symptoms weeks/months after potential exposure
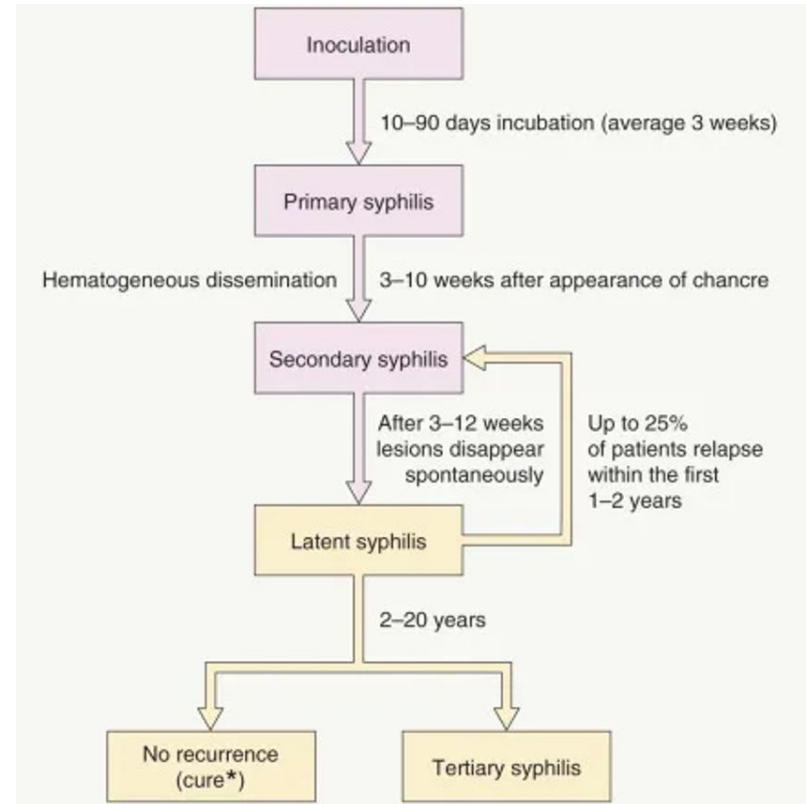
Diagnosis/Treatment

- diagnosis: symptoms, lab test
- treatment: penicillin

Risk Factors

- risky sexual behaviors
- mother having syphilis or herpes

Populations at Risk

- sexual minorities
- racial/ethnic minorities
- men

Inoculation

↓ 10–90 days incubation (average 3 weeks)

Primary syphilis

Hematogeneous dissemination → 3–10 weeks after appearance of chancre

Secondary syphilis

After 3–12 weeks lesions disappear spontaneously

Up to 25% of patients relapse within the first 1–2 years

Latent syphilis

2–20 years

No recurrence (cure*) Tertiary syphilis

Measures of Disease Occurrence

Morbidity, Mortality

Types of Measures

Prevalence

Incidence

Measures of Disease

- Measures of disease allow us to quantify disease in terms of morbidity and mortality
- **Morbidity**: Those living with disease
- **Mortality**: Those who have died from disease

Morbidity

- "Any departure, subjective or objective, from a state of physiological or psychological well-being"
- Comorbidities: Multiple illnesses experienced at the same time

Mortality

- Every human being will experience death
- Important to note *when* and *how* death is experienced
  - Can death be delayed?
- Expressing mortality in quantitative terms can pinpoint differences in the **risk** of dying from a disease:
  - Between people in different geographic regions
  - Between subgroups in the population
- **Mortality rates** can serve as measures of disease severity and can help us determine whether the treatment for a disease has become more effective over time

General Types of Measures

- **Counts**: Absolute number of occurrences in a snapshot of time

  - Useful for public health efforts, burden of disease

  - EX: 12 individuals with colds in class on Friday

- **Proportions**: What fraction of population affected by disease?

  - Numerator is always included in the denominator

  - EX: 12% of class of 100 students with a cold

- **Rates**: How fast the disease occurs in population

  - Time included in the denominator

  - EX: 200 cases per 100,000 population per year

- **Ratios**: One number or rate relative to another

  - EX: 5 male cases of breast cancer to 40 females of breast cancer, i.e ratio of 5:40
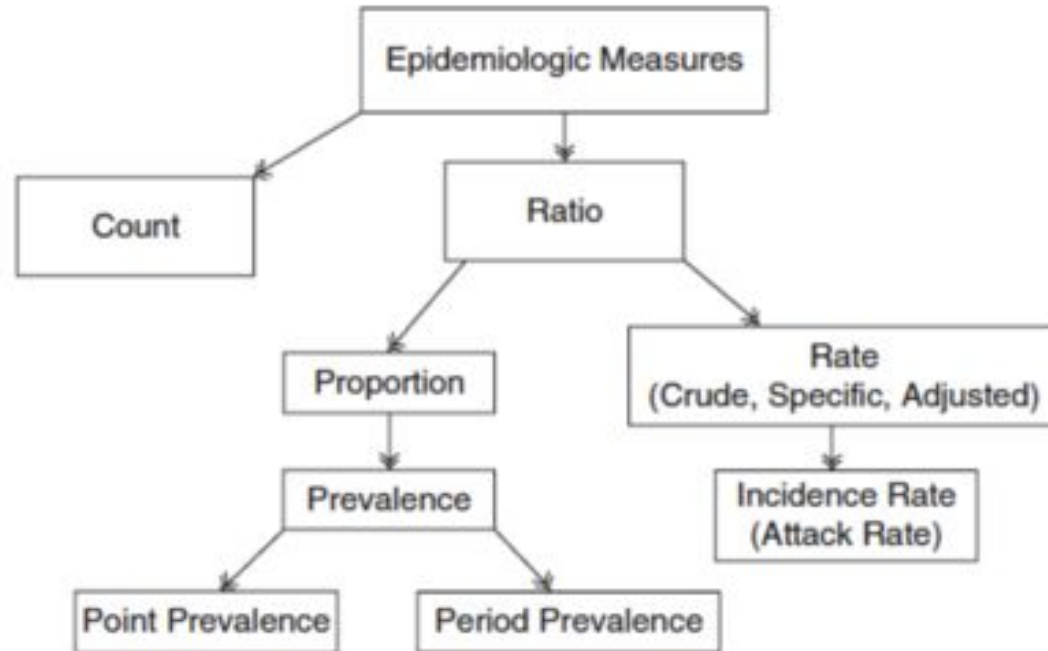
General Types of Measures

FIGURE 3–1   Overview of epidemiologic measures.

Measures of Morbidity

Prevalence

Incidence

Relating Prevalence and Incidence

Prevalence

- Number of **existing** cases at a certain time

- 2 measures: Point Prevalence, Period Prevalence

Point Prevalence	Period Prevalence
$\dfrac{\text{\# existing cases of disease at time t}}{\text{\# in total population at time t}}$	$\dfrac{\text{\# indiv cases existing or occurring at any point btwn t1 and t2}}{\text{\# pop'n at midpoint btwn t1 and t2}}$

Incidence

- Number of **new cases** of disease in a certain time period

Cumulative incidence	Incidence density, or Incidence rate
$$\frac{\text{\# new cases of disease}}{\text{\# of ppl "at risk" in pop'n at the beginning of the period}}$$	$$\frac{\text{\# new cases of disease}}{\text{Total person} - \text{time of observation}}$$
- Assumes that everyone has been followed for the specified time period. - Usually reserved for fixed populations. - Denominator is people **at risk** at the beginning of the period; that's why it's a proportion! - Literally the cumulative number of incident cases expressed as a proportion.	- Does NOT assume that everyone has been followed for the specified time period. - Person-time is total time under observation for all individuals. - Individuals do not have to be observed for same amount of time. - Person-time ends when individual becomes diseased.

Incidence and Prevalence are related

Prevalence ≅ Incidence x Duration

epidemiologist's bathtub

Measures of Mortality

Overview of measures

Measures of Mortality: Overview

Measure	Numerator	Denominator	Type
Mortality rate	# of deaths from all causes	Total population	Proportion
Cause-specific mortality rate	# of deaths from a specific disease/condition	Total population	Proportion
Proportionate mortality	# of deaths from a specific disease/condition	Total number of deaths	Proportion
Case-fatality rate	# of deaths from a specific disease/condition	Total number of persons with specific disease/condition	Proportion

These measures can be considered a type of incidence but instead of the occurrence of new cases of disease or morbidity, **it is the occurrence of new cases of death.** Incidence measures the risk of morbidity and mortality measures the risk of death.

Study Designs

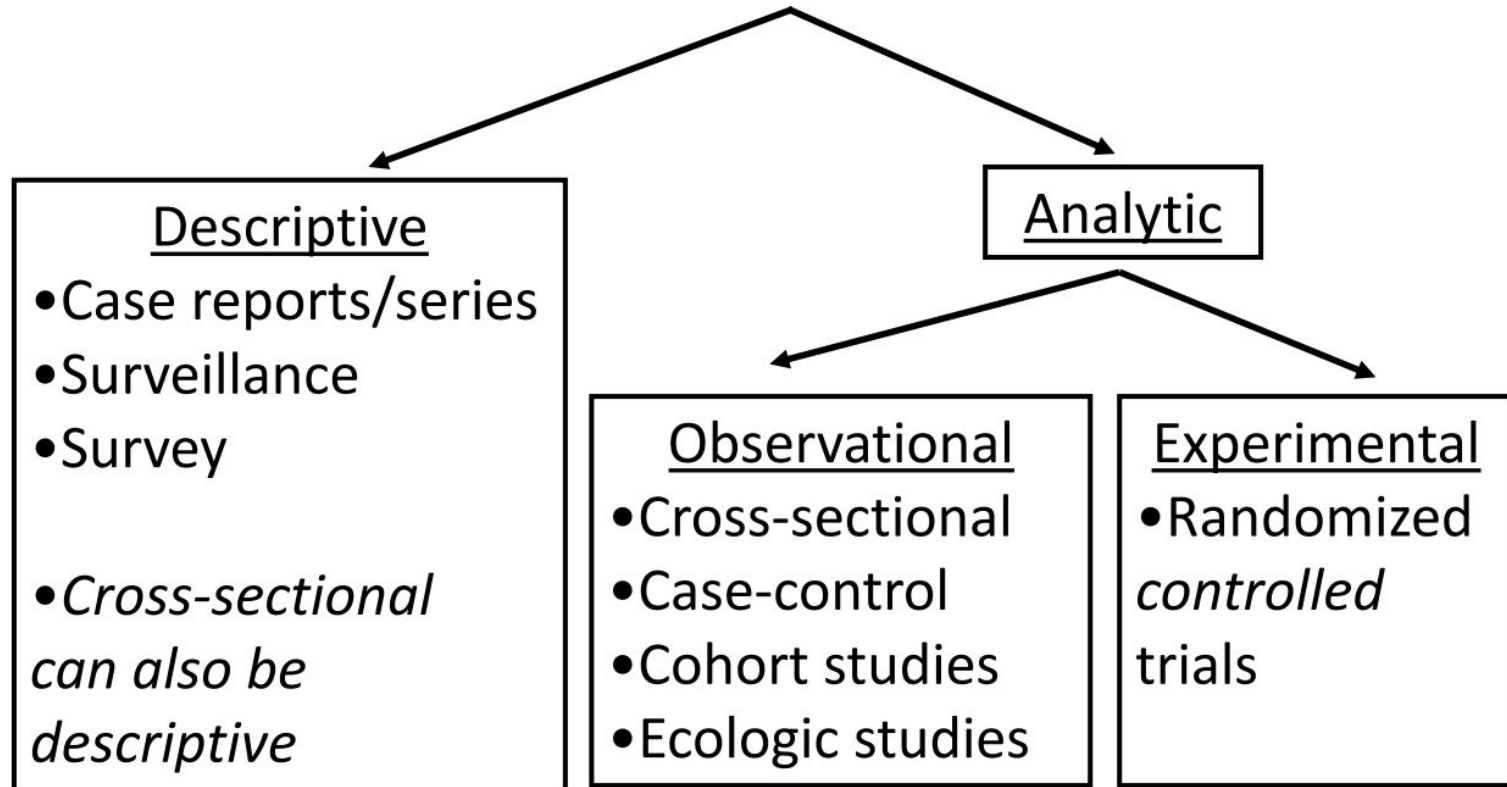Hierarchy of Epidemiology Studies

Components of Studies

Ecological

Cross-Sectional (Prevalence)

Cohort

Case-Control

Hierarchy of Epidemiological Studies

Descriptive vs. Analytic Studies

- **Descriptive Studies**: Count and describe outcome
  - (who, when, where, how many?)
  - EX: Surveys, disease surveillance, cross-sectional, ecologic
- **Analytic Studies**: analyze factors that may affect outcome (how and why?)
  - EX: ecologic, cross-sectional, case-control, cohort, experimental studies
  - Analytic studies attempt to identify **risk factors** for disease

???

Smoking ➡ **Asthma**
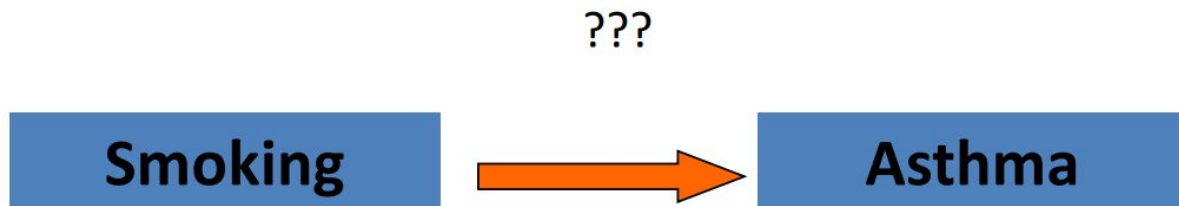
Components of Analytic Studies

- A study question

- A defined study group

- Information on study outcome

- Information on study exposure(s)

- Assessment of the association between exposure and outcome

- Conclusion

Common Terminology

|	
Study group	sample, study population, cases/controls
Information	characteristics, factors, variables, parameters
Outcome	disease, endpoint, dependent variable
Exposure	predictor, independent variable, risk factor
Other exposures	covariables, covariates, confounders, cofactors
Association	relationship, correlation

Example: Breast Cancer

- Define the Question: Is a particular diet associated with an increased number of new breast cancer cases?

- Select study group (individuals or group)

- Outcome: diagnosed with breast cancer: Yes/No

- Exposure: one example could be fruit and vegetable intake

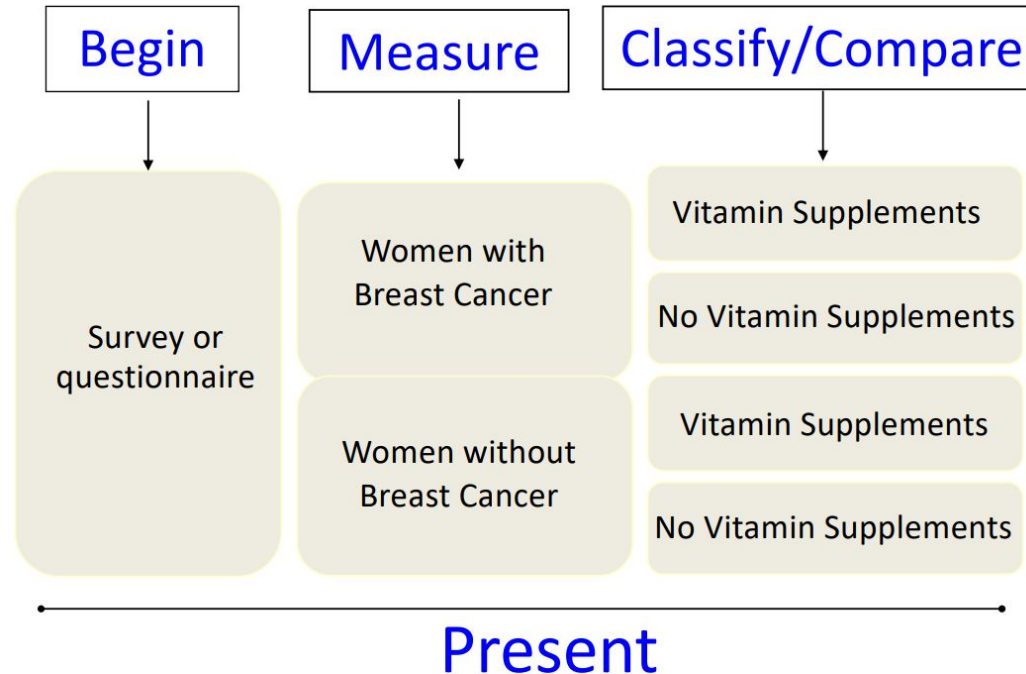- Are exposure and outcome related? (correlated, associated)

Ecological Studies

- Studies in which the unit of analysis are large, aggregated groups or populations
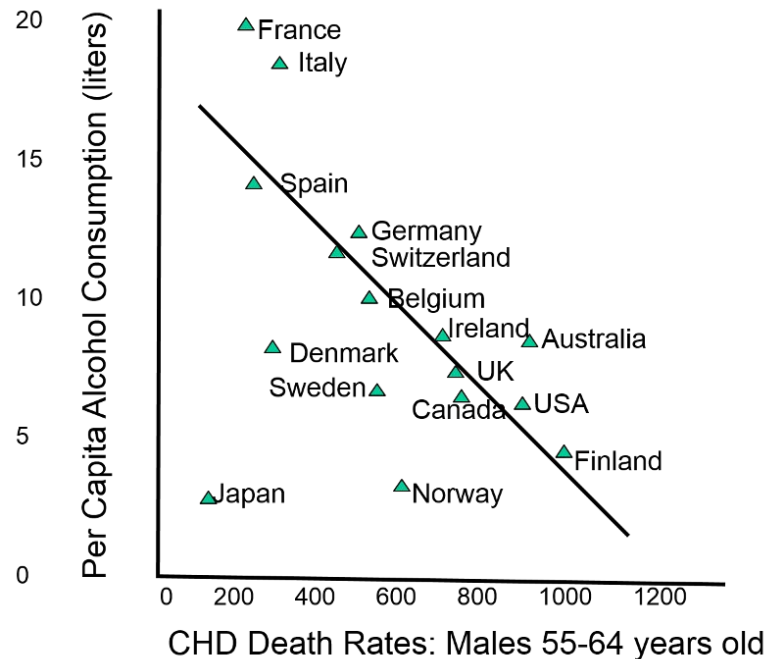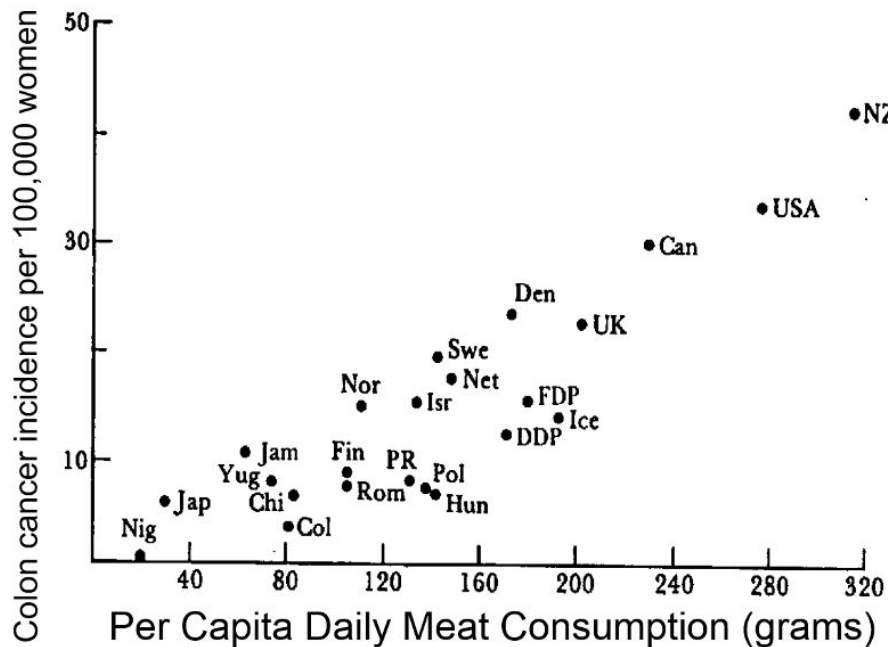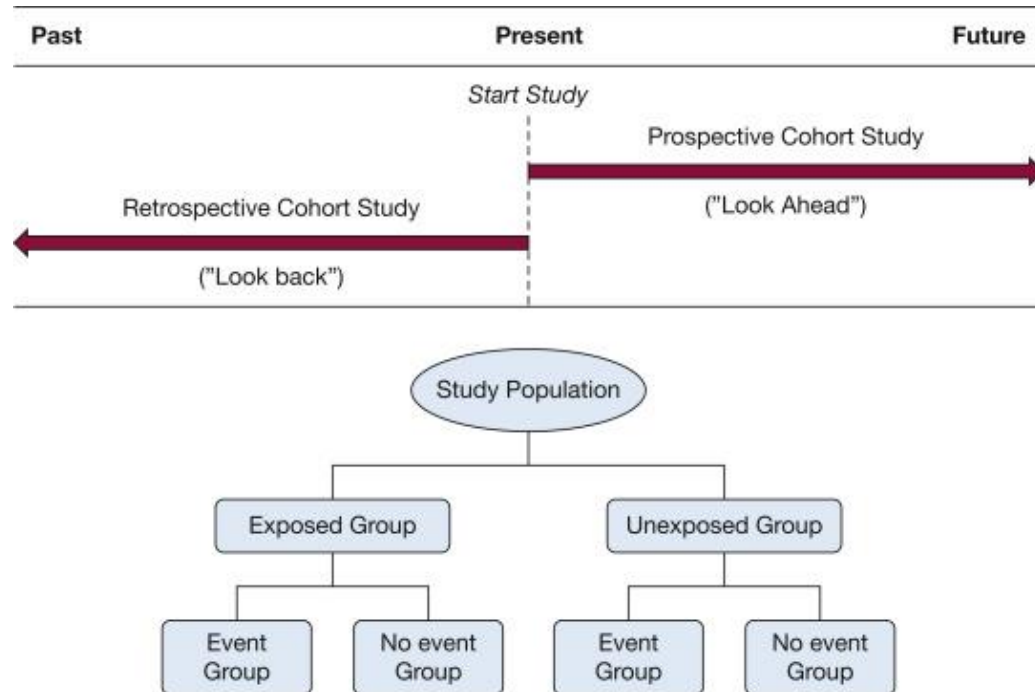
Cross-Sectional (Prevalence) Study

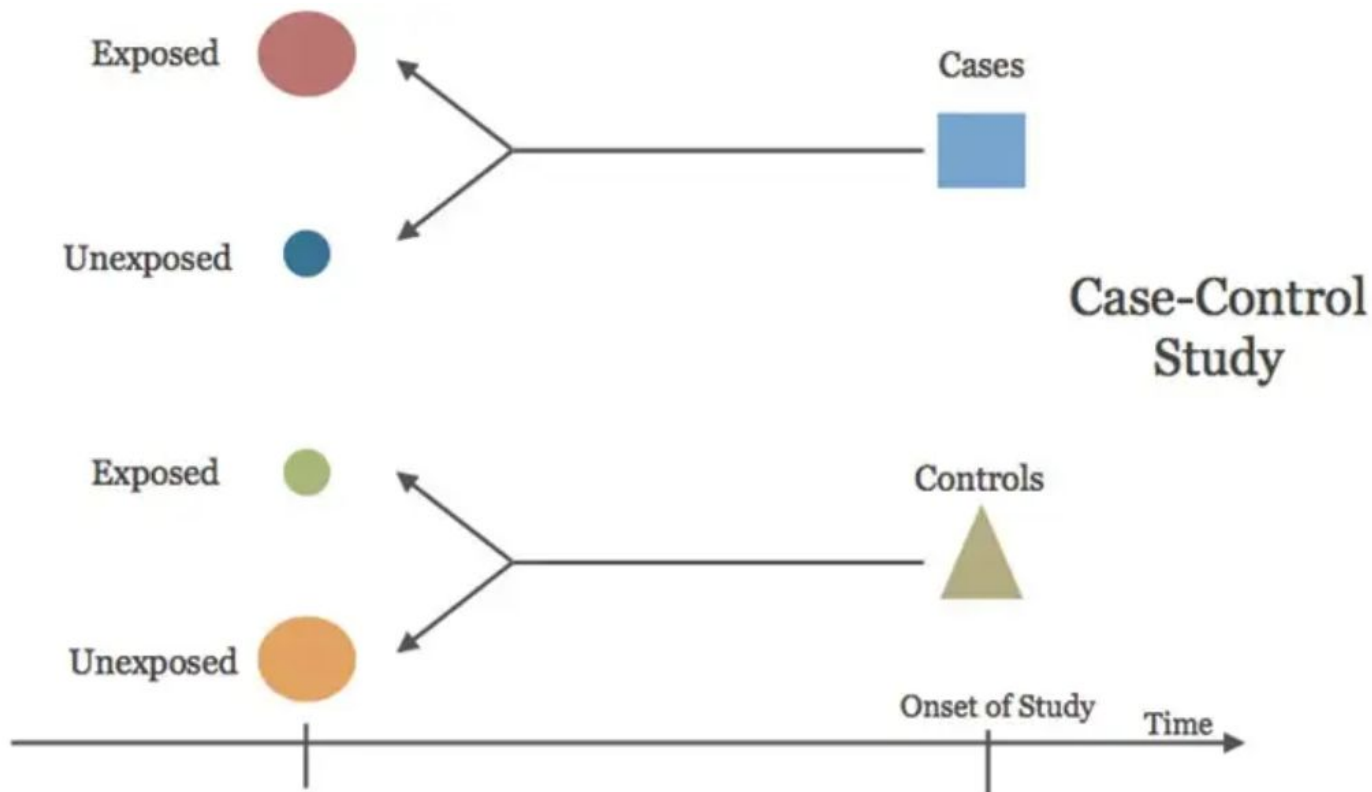- Studies that record population information at one snapshot in time

Cohort Studies

- Retrospective, Prospective

Case-Control Studies

CROSS SECTIONAL STUDY — Individuals examined at a single time point

CASE CONTROL STUDY

Factors present or absent ← Individuals with a particular outcome

Factors present or absent ← Individuals without a particular outcome

COHORT STUDY (RCT)

Individuals exposed to particular factors → Particular outcome present or absent

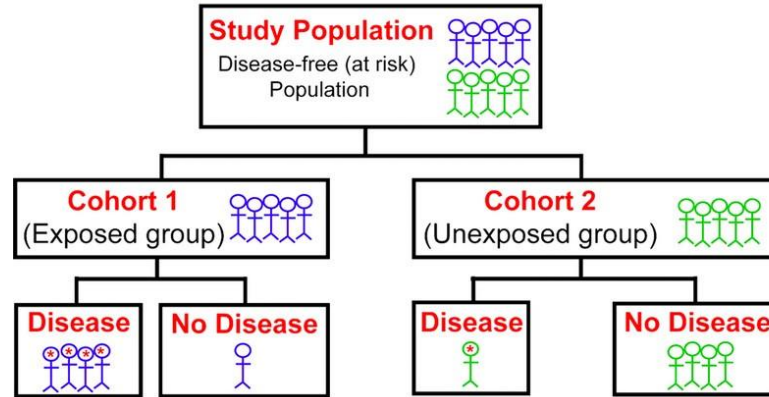Individuals not exposed to particular factors → Particular outcome present or absent

TIME →
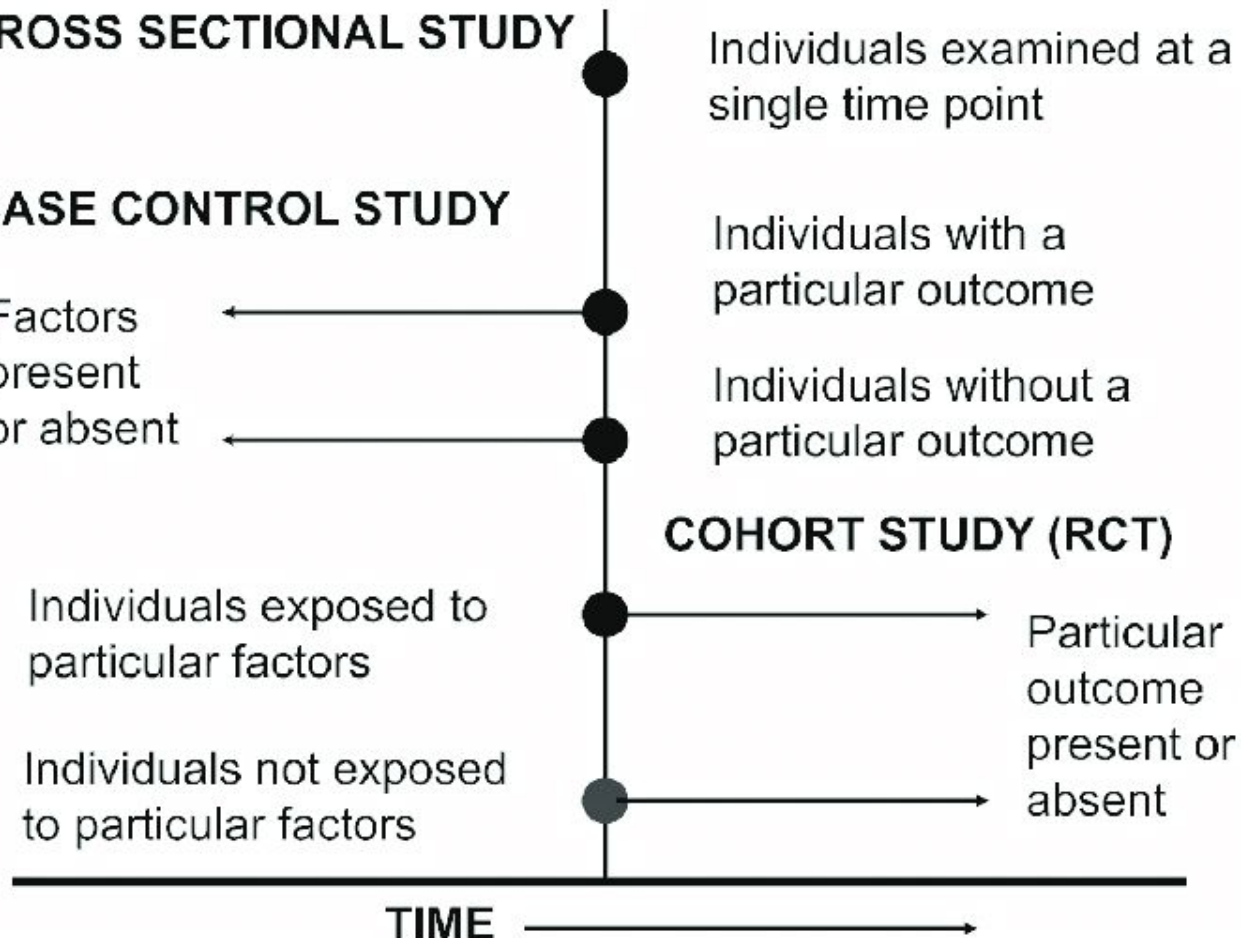
Cohort vs Case-Control

COHORT STUDY

Study Population
Disease-free (at risk) Population

Cohort 1
(Exposed group)

Cohort 2
(Unexposed group)

Disease | **No Disease** | **Disease** | **No Disease**

1. Identify exposed and unexposed cohort groups.

   - Unexposed
   - Exposed
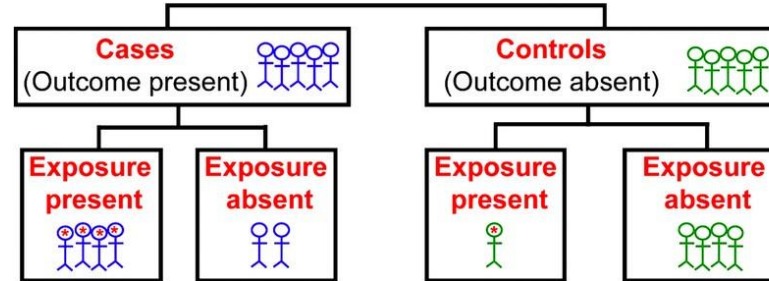   - * Diseased status

2a. PROSPECTIVE STUDY
   -During follow-up period, identify diseased subjects (incident cases).

2b. RETROSPECTIVE STUDY
   -Identify diseased subjects by interview or written records.

3. Analyze differences (i.e. incidence or relative risk) among those exposed (cohort 1) and those unexposed (cohort 2).

CASE-CONTROL STUDY

Cases
(Outcome present)

Controls
(Outcome absent)

Exposure present | **Exposure absent** | **Exposure present** | **Exposure absent**
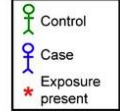
1. Identify cases.

   - Control
   - Case
   - * Exposure present

2. Select controls, which may be matched to cases.

3. Measure exposure or risk factors of interest.

4. Compare the presence or absence of exposure in cases and controls.

Exercise: Interventions

Determine the study design type.

Questionnaires were mailed to every 10th person listed in the city telephone directory. Each person was asked to list age, sex, smoking habits, and respiratory symptoms during the preceding seven days. About 20% of the questionnaires were completed and returned. About 10% of respondents reported having upper respiratory symptoms.

Exercise: Interventions

Determine the study design type.

Questionnaires were mailed to every 10th person listed in the city telephone directory. Each person was asked to list age, sex, smoking habits, and respiratory symptoms during the preceding seven days. About 20% of the questionnaires were completed and returned. About 10% of respondents reported having upper respiratory symptoms.

Cross-sectional (prevalence)

Exercise: Interventions

Determine the study design type.

A random sample of middle-aged sedentary adults were selected from four census tracts, and each person was examined for coronary artery disease (CAD). All persons without disease were randomly assigned to either a two-year program of aerobic exercise or a two-year arthritis-prevention non-aerobic exercise program. Both groups were observed semiannually for incidence of CAD.

Exercise: Interventions

Determine the study design type.

A random sample of middle-aged sedentary adults were selected from four census tracts, and each person was examined for coronary artery disease (CAD). All persons without disease were randomly assigned to either a two-year program of aerobic exercise or a two-year arthritis-prevention non-aerobic exercise program. Both groups were observed semiannually for incidence of CAD.

Randomized Controlled Trial (RCT)

Determine the study design type.

Patients admitted for carcinoma of the stomach and patients without a diagnosis of cancer are interviewed about their chewing tobacco history to assess the possible association of chewing tobacco and gastric cancer.

Exercise: Interventions

Determine the study design type.

Patients admitted for carcinoma of the stomach and patients without a diagnosis of cancer are interviewed about their chewing tobacco history to assess the possible association of chewing tobacco and gastric cancer.

Case-control study

Exercise: Interventions

Determine the study design type.

1,500 employees of a major aircraft company were initially examined in 1951 and were classified by diagnostic criteria for coronary artery disease (CAD). New cases of CAD have been identified by examinations every three years and through death certificates. Attack rates in different subgroups have been computed.

Exercise: Interventions

Determine the study design type.

1,500 employees of a major aircraft company were initially examined in 1951 and were classified by diagnostic criteria for coronary artery disease (CAD). New cases of CAD have been identified by examinations every three years and through death certificates. Attack rates in different subgroups have been computed.

Prospective cohort study

Exercise: Interventions