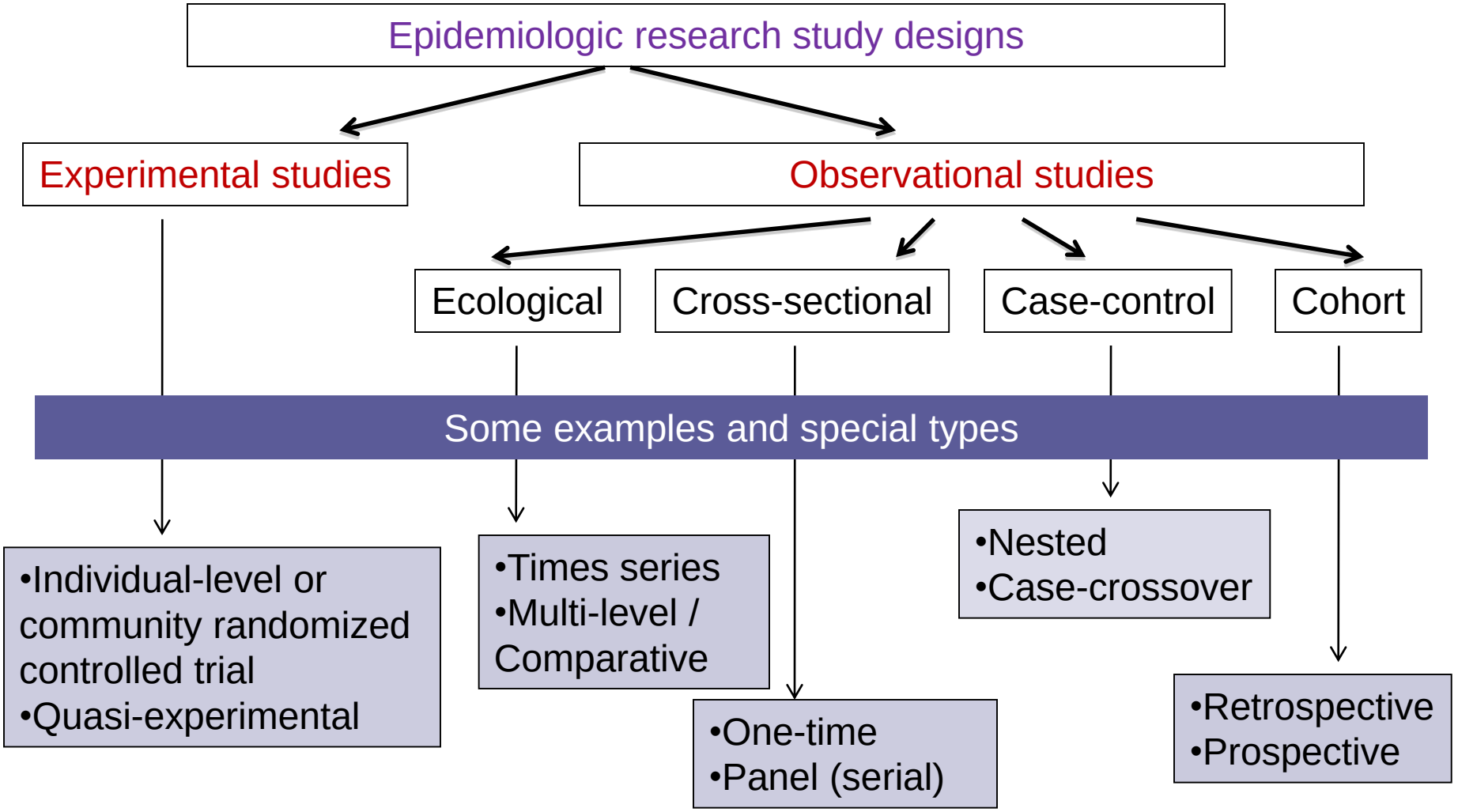
STUDY DESIGN : ECOLOGICAL & CROSS SECTIONAL

Assoc. Prof. Dr Azmawati Mohammed Nawi
Epidemiology & Statistic Unit,
Dept. of Community Health, UKM

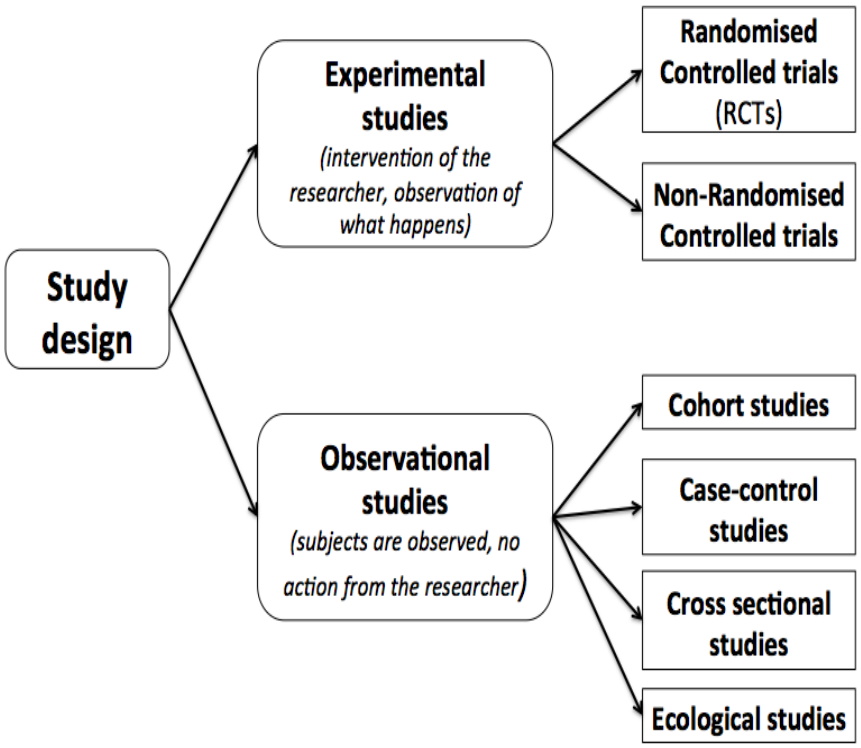
Overview for Today

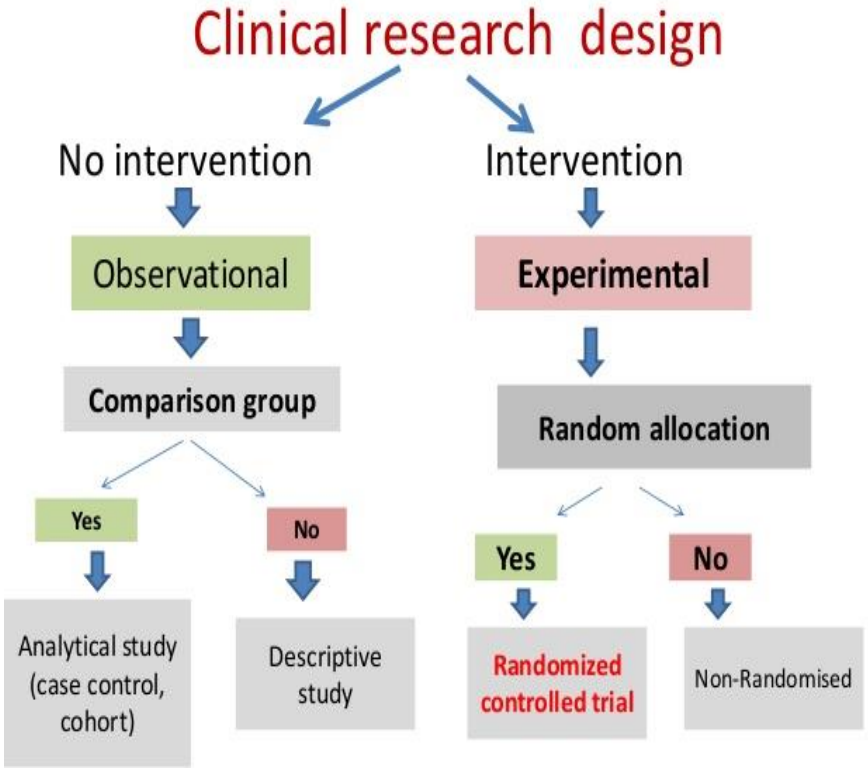
- Overview of Types of Study Designs
- Ecological studies
  - Definitions, uses, strengths and limitations
  - Examples
- Cross-sectional studies
  - Definitions, uses, strengths and limitations
  - Examples

Overview : Main Types of Epidemiologic Studies



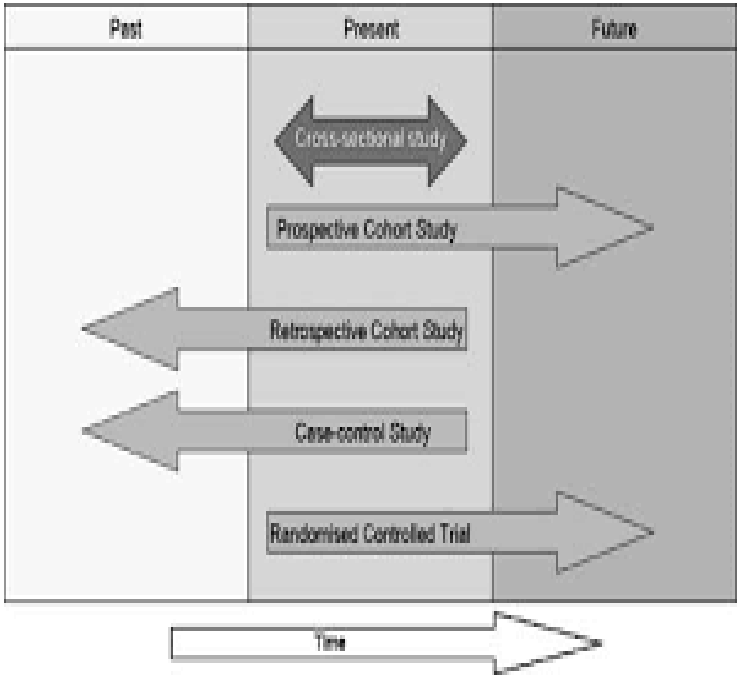
Overview : Main Types of Epidemiologic Studies



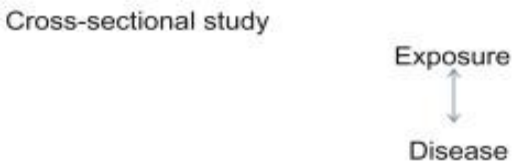


Commonly Used Definitions

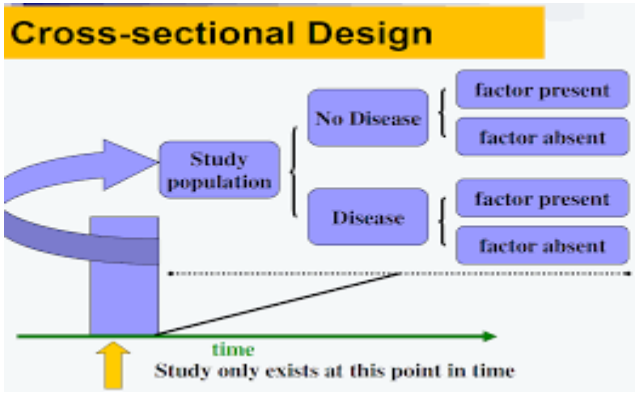
- Observational
  - Studies of disease burden or etiology, investigator passively observes nature
- Ecological
  - Examines the relationship between exposure and disease with population-level rather than individual-level data
- Cross-sectional
  - Examines disease burden or relationship between exposure and disease at a single point in time
- Case-control
  - Examines exposure-disease relationship by enrolling cases (with disease) and controls (without disease) and comparing exposure history
- Cohort
  - Examines exposure-disease relationship by enrolling individuals without disease, assessing exposures, and following over time for disease occurrence
- Experimental
  - Studies in which the investigator actively manipulates which groups receive a preventive or treatment intervention of interest, and then follows over time

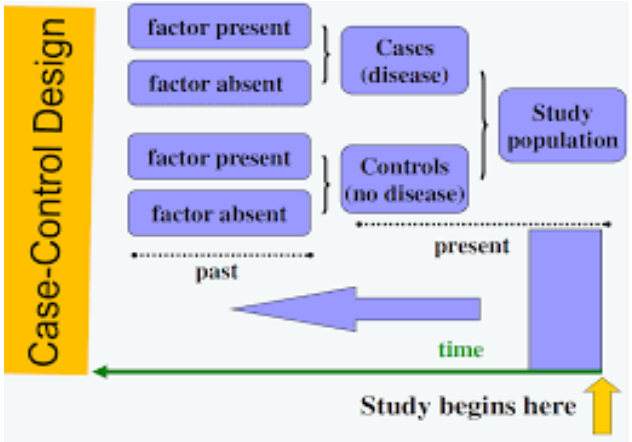

Past
Present
Future
Cross-sectional study
Prospective Cohort Study
Retrospective Cohort Study
Case-control Study
Randomised Controlled Trial
Time

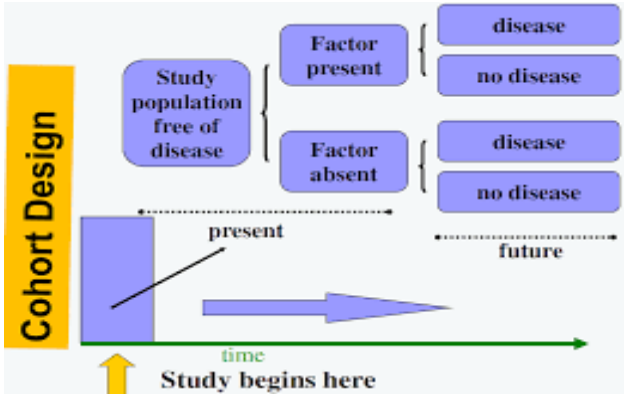

Case–control study
Exposure
Disease


Cross-sectional study
Exposure
Disease

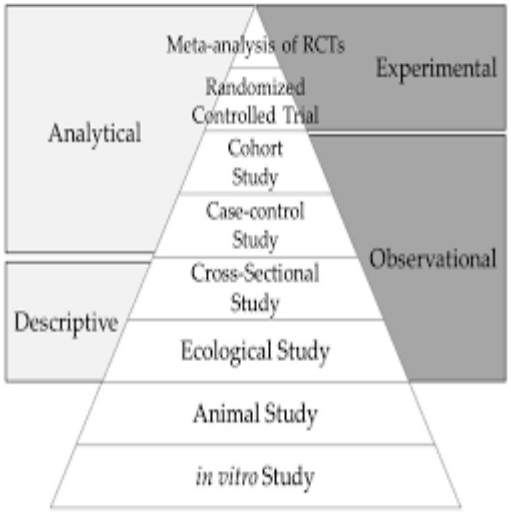

Cohort study
Exposure
Disease
Time


Cross-sectional Design
Study population
No Disease
Disease
factor present
factor absent
factor present
factor absent
time
Study only exists at this point in time


Case-Control Design
factor present
factor absent
factor present
factor absent
Cases (disease)
Controls (no disease)
Study population
past
present
time
Study begins here


Cohort Design
Study population free of disease
Factor present
Factor absent
disease
no disease
disease
no disease
present
future
time
Study begins here

ECOLOGICAL STUDY



ECOLOGICAL STUDY

- Definition: observational study in which group being unit of observation instead of individual
- The group could be defined by
  – time (calendar period, birth cohort),
  – geography (country, city),
  – socio-demographic characteristics (ethnicity, religion).
- Types:
  - Time series / trend study : examine an exposure-disease relationship over several time points in a single group
  - Comparative : Multiple group comparison study (e.g. geographical correlation study) ; examine an exposure-disease relationship between groups at one point in time

TYPE OF ECOLOGIC STUDY DESIGNS

Based upon Groups of an Ecologic study	Based upon the Method of exposure measurement in Ecological study	
a. by place (multiple-group design)	Exploratory	Analytical/Etiological
b. by time (time-trend design)	Exploratory	Analytical/Etiological
c. by combination of place and time (mixed design)	Exploratory	Analytical/Etiological

Hypothetically…..

- Determine the proportion of people in a country who smoke and the annual death rate from lung cancer in that country for several countries and compare.

- In a single country, determine how the proportion of people who smoke has changed over time, and how the annual death rate from lung cancer has changed over time.

Ecological Example: Doll & Hill, 1950

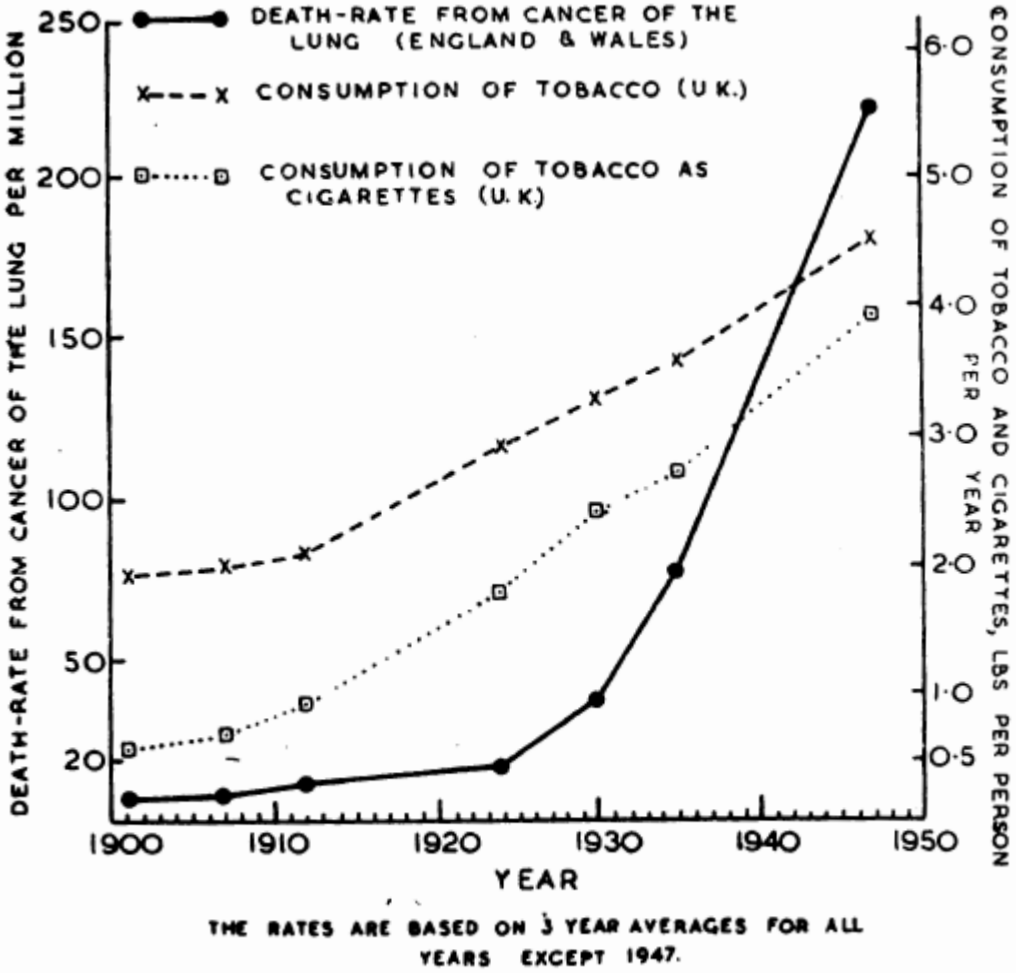


FIG. 2.—Death rate from cancer of the lung and rate of consumption of tobacco and cigarettes.

Ecological study: comparison of cancer rates in municipalities with and without chlorinated drinking water

Chlorination of Drinking Water and Cancer Mortality in Taiwan

Chun-Yuh Yang,* Hui-Fen Chiu,† Ming-Fen Cheng,‡ and Shang-Shyue Tsai‡

*School of Public Health, Kaohsiung Medical College, Kaohsiung, Taiwan; †Department of Pharmacology, Kaohsiung Medical College, Kaohsiung, Taiwan; and ‡School of Public Health, Kaohsiung Medical College, Kaohsiung, Taiwan

Received July 30, 1997

Chlorination has been the major strategy for disinfection of drinking water in Taiwan. An ecologic epidemiological study design was used to examine whether chlorination of drinking water was associated with cancer risks. A "chlorinating municipality" (CHM) was defined as one in which more than 90% of the municipality population was served by the chlorinated water while an "nonchlorinating municipality" (NCHM) was one in which less than 5% of the municipality population was served by chlorinated water. Age-adjusted mortality rates for cancer during 1982–1991 among the 14 CHMs were compared to rates among the 14 matched NCHMs with similar urbanization level and sociodemographic characteristics. The results of this study suggest a positive association between consumption of chlorinating drinking water and cancer of the rectum, lung, bladder, and kidney. Although these findings must be interpreted with caution because of limitations in the ecological study design

and Comstock, 1981; Young *et al.*, 1981,1987; Bean *et al.*, 1982; Gottlieb *et al.*, 1982; Lawrence *et al.*, 1984; Carpenter and Beresford, 1986; Cech *et al.*, 1987; Zierler *et al.*, 1988; Flaten, 1992; Morris *et al.*, 1992; McGeehin *et al.*, 1993). These studies consider a wide range of populations and regions but have been mainly carried out in the U.S. Most studies have shown positive associations between chlorinated drinking water and colorectal and bladder cancer. This has been attributed to trihalomethanes (THMs), a carcinogenic organic halogenated byproduct of water chlorination (Reuber, 1979; Dunnick and Melnick, 1993; IARC, 1987).

The present study was carried out because few epidemiological studies have been conducted outside the U.S. (Carpenter and Beresford, 1986; Flaten, 1992). The study reported here was designed to explore further the association between cancer mortality and the use of chlorinated water.

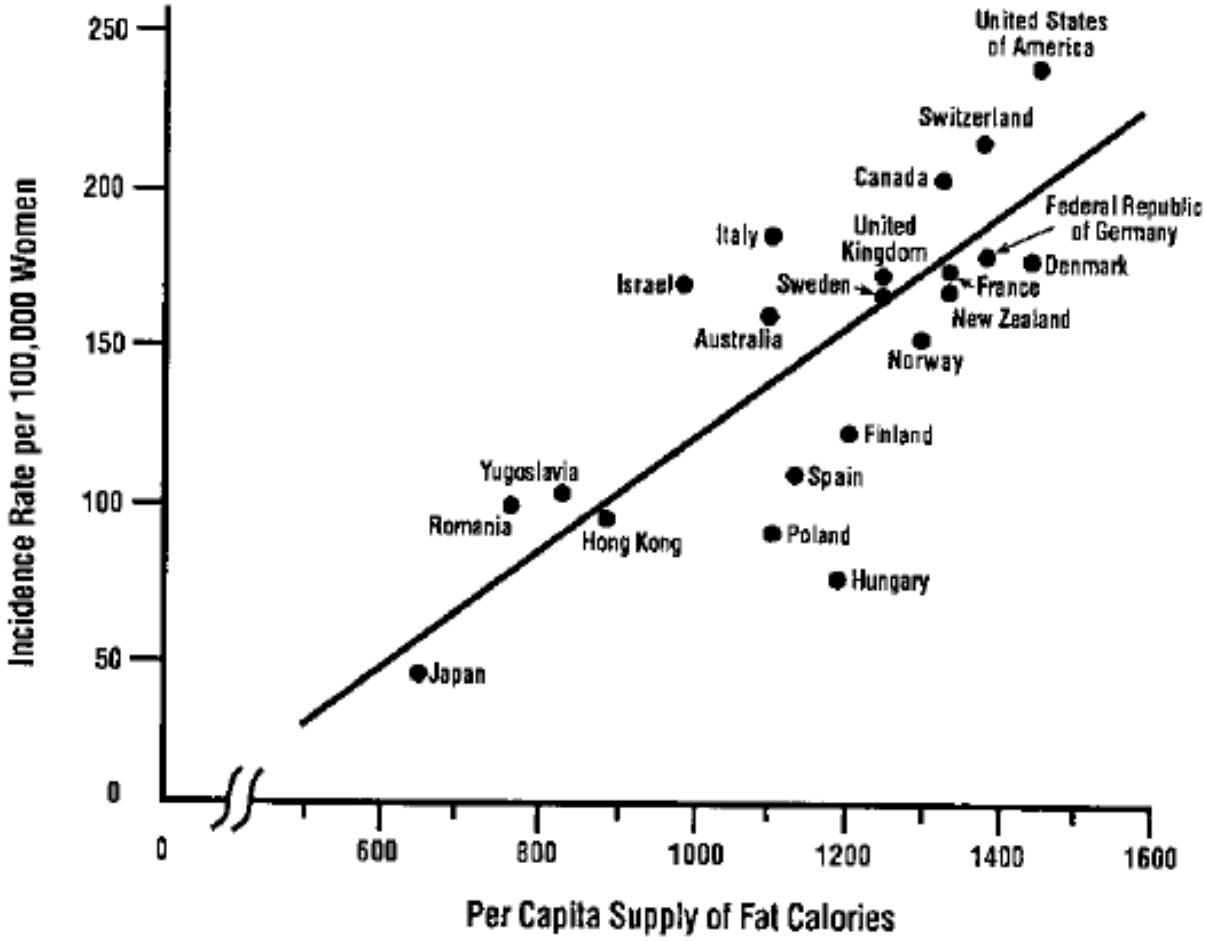


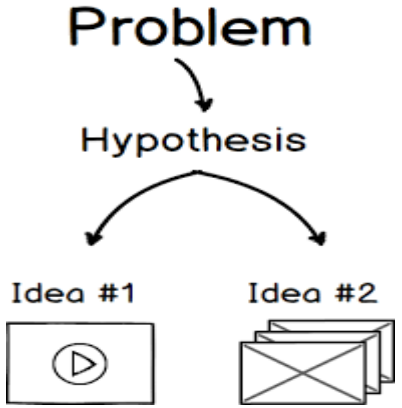
Figure 13–4. Correlation between dietary fat intake and breast cancer by country. (From Prentice RL, Kakar F, Hursting S, et al: Aspects of the rationale for the Women's Health Trial. J Natl Cancer Inst 80:802–814, 1988.)

Study design: Ecological & Cross sectional

Azmawati

ADVANTAGES

- Can generate interesting hypotheses
- May be useful for generating leads/hypothesis for pursuit in more definitive studies
- Can rely on routinely collected data (census information, vital statistics, disease registry data)
- Are relatively quick and cheap



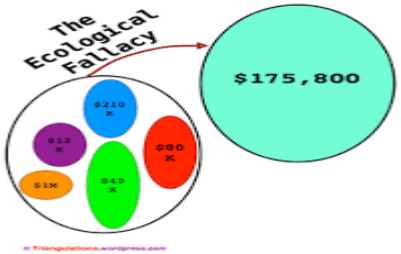
DISADVANTAGES

- Associations may not be causal
  - Could be due to a third factor or intermediate variable
  - Limited for testing etiologic hypotheses
- May not have data on confounders
  - Especially when using existing data sources
- The "**Ecological Fallacy**"
  - an association observed between variables at the aggregate level may not necessarily represent the association that exists at the individual level
  - Do not know joint distribution of risk and outcome
  - Existing data sources are often compiled independently

THE ECOLOGICAL FALLACY

- Common meaning: The tendency of ecological studies to imprecisely reflect reality as documented in individual studies
- Statistical meaning: The tendency of correlation coefficients to be larger when an association is assessed at the group level than when it is assessed at the individual level

EXAMPLE OF ECOLOGICAL FALLACY



- Study of the rate of coronary heart disease in the capital cities of the world vs average income.
  - Within the cities studied, coronary heart disease is higher in the richer cities than in the poorer ones.
  - General conclusion might be ....being rich increases your risk of heart disease.
  - However, individual studies in developed and industrialised cities such as London, Washington and Stockholm, poor people found to have higher CHD rates than the rich.
- The ecological fallacy is usually interpreted as a major weakness of ecological analyses

Analysis of Ecological Data

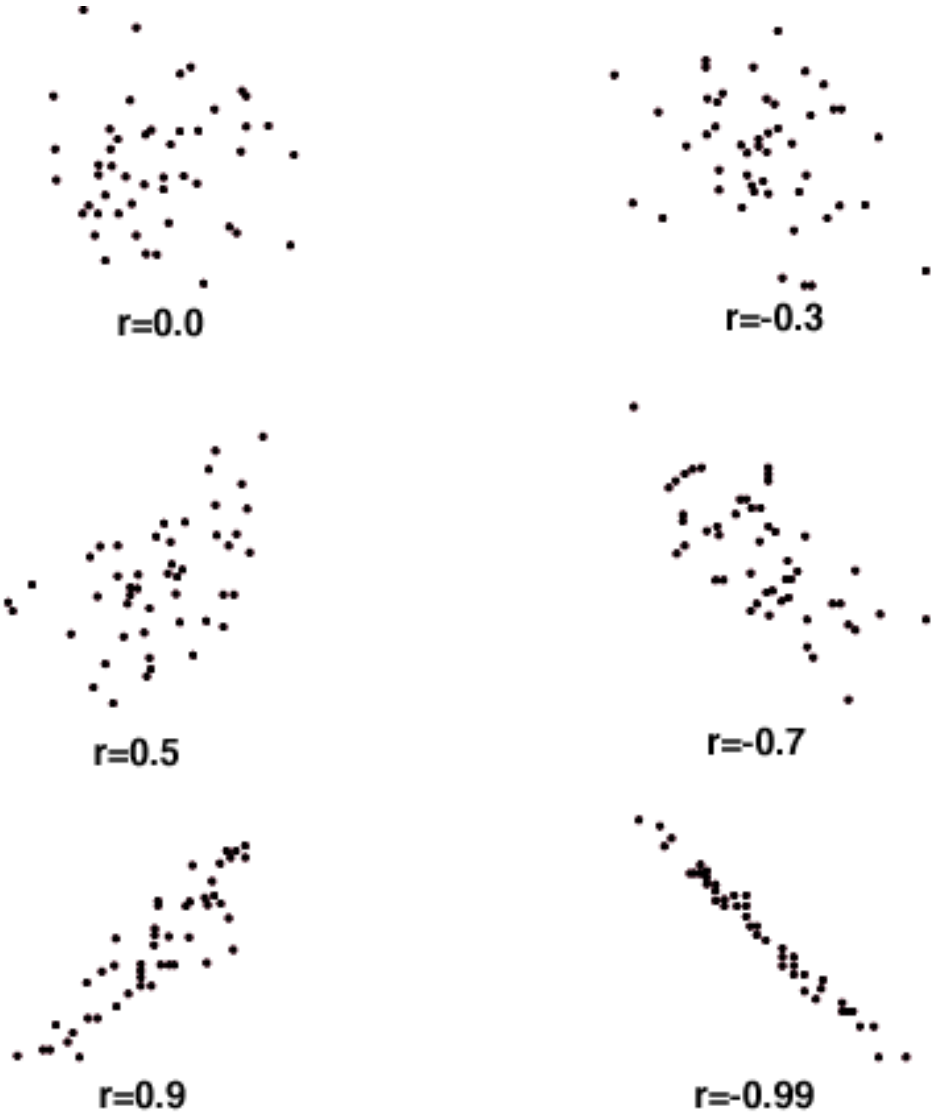
- Unit of analysis is population or group, not the individual
- Slightly different than other study designs
  - Cannot fill in 2x2 table with individuals (units) in cells
- Instead, determine the relationship between the exposure variables (independent variables, x_1,x_2, ...) and the disease variables (dependent variables, y_1,y_2, ...)
  - Plot data in two dimensions
  - Estimate r
  - Estimate r^2

  (Regression analysis)

r (Correlation Coefficient)

- A measure of association that indicates the degree to which 2 variables have a linear relationship
- Can vary between +1 and –1
- r > 0
  - A group with a high/low value for one variable will likely have a high/low value for the other
- r < 0
  - A group with a high/low value for one variable will likely have a low/high value for the other
- r = 0
  - The variables are independent
- Larger values → stronger correlation

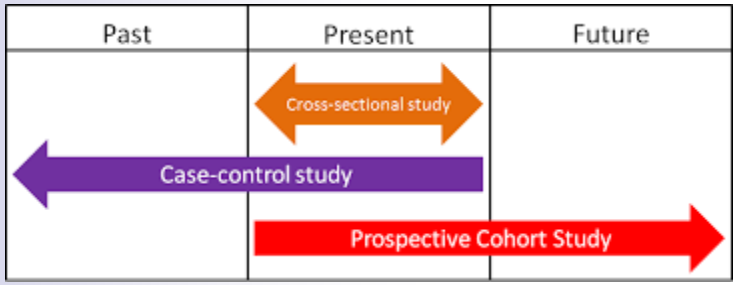
Some Examples of x-y Plots

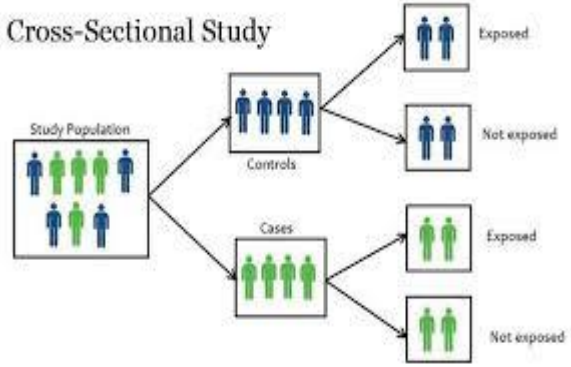


R^2 (Coefficient of Determination)

- The square of the correlation coefficient, R^2, represents the proportion of the regression of Y on X that can be attributable to X
- Proportion of variance in y that is accounted for by a linear fit of x to y
  - When r is 0.20, R^2 is 4%
  - When r is 0.70, then R^2 is about 50%
  - When r is 0.90, then R^2 is about 80%

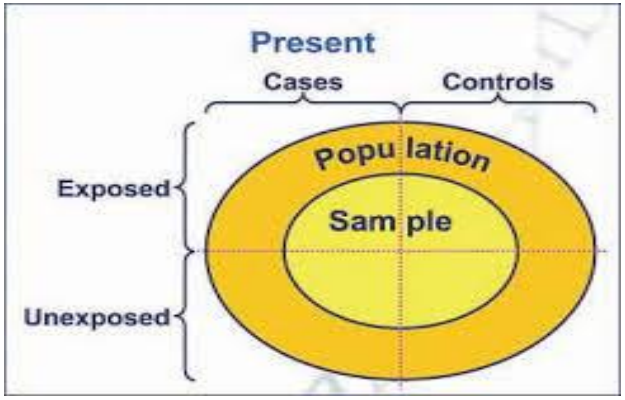
CROSS SECTIONAL STUDY





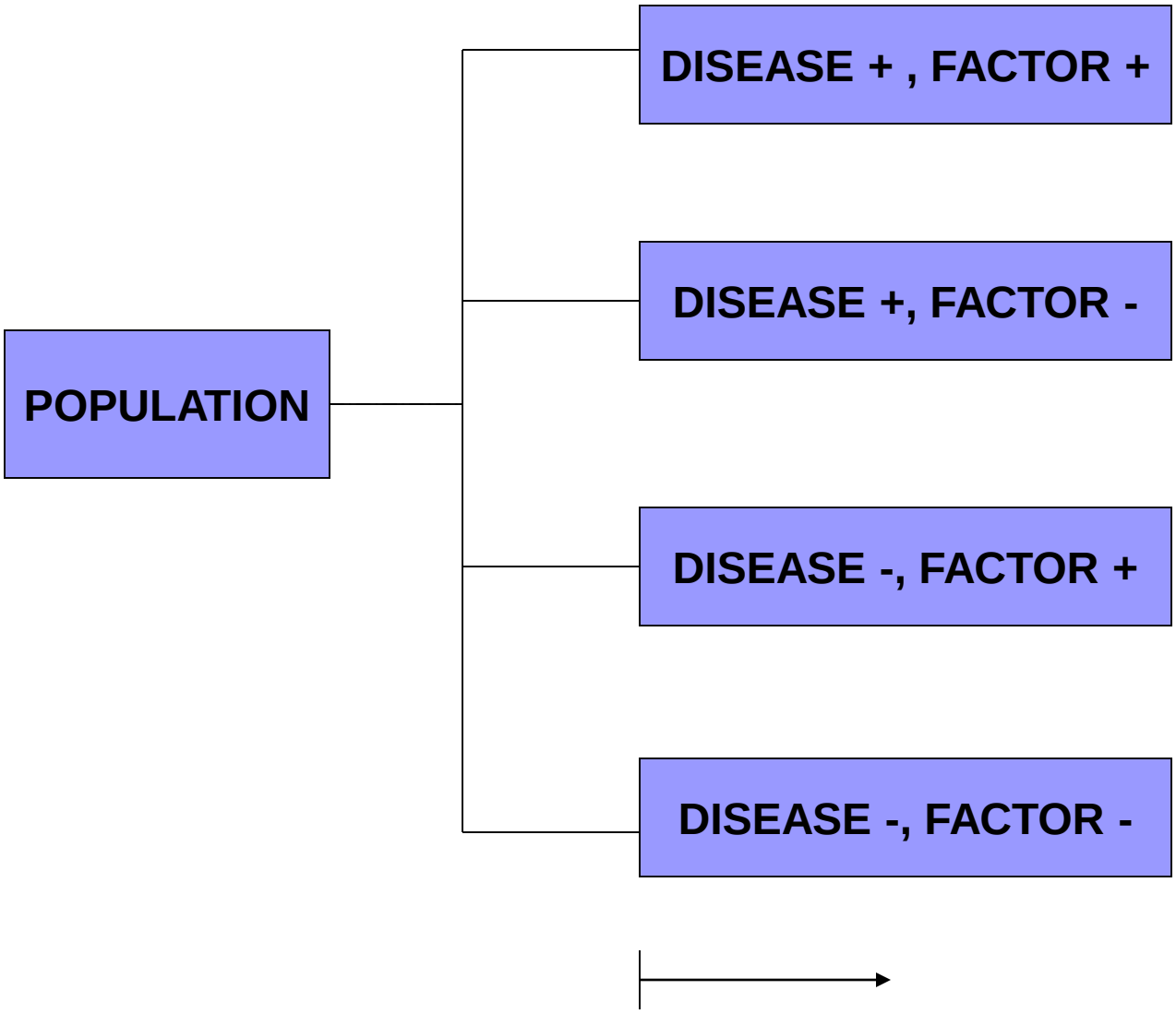
CROSS SECTIONAL STUDY

- Also known as cross sectional surveys, epidemiologic studies, and prevalence studies.
- Cross sectional is more appropriate because of description of the time line
- Examines the relationship between diseases and other variable of interest as they exist in a defined population at one particular time

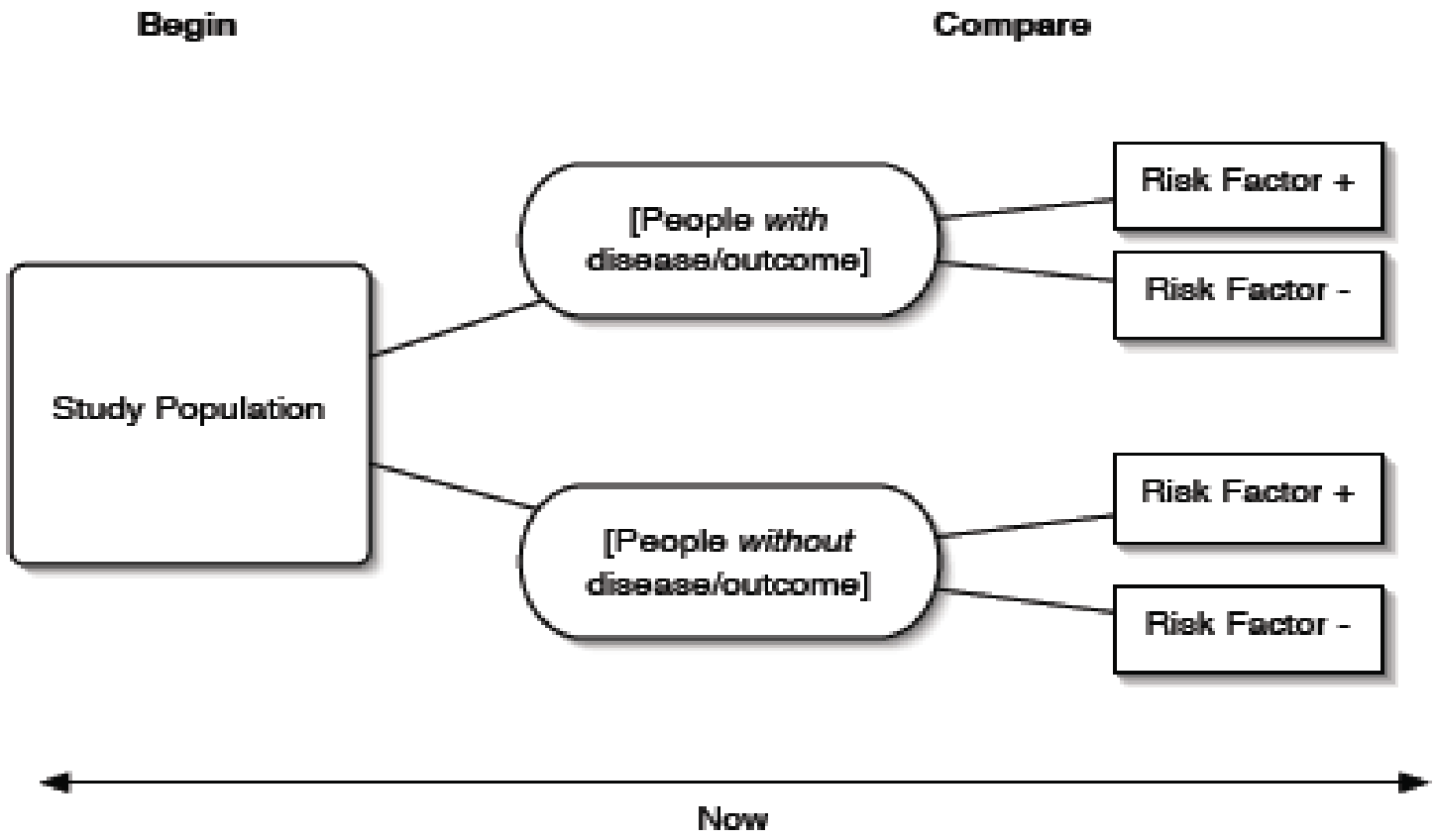


CROSS SECTIONAL STUDY

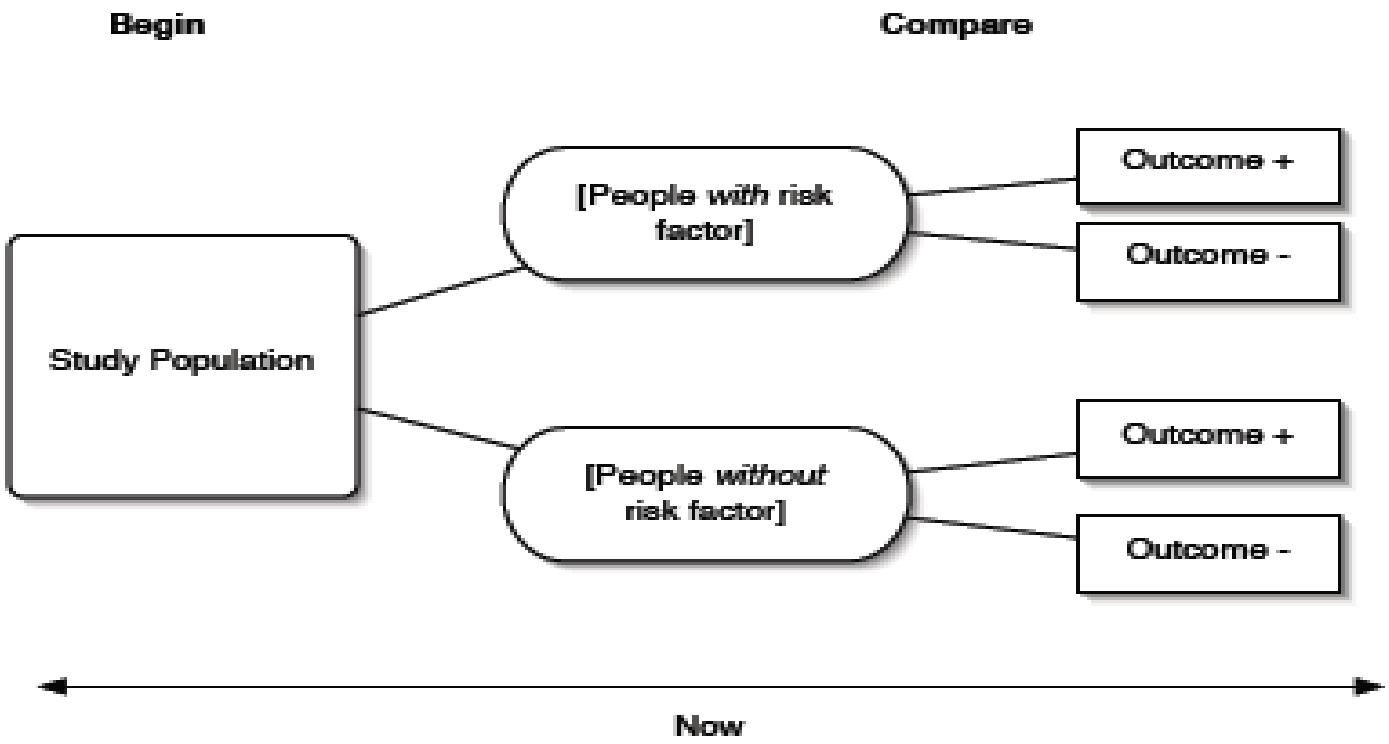
- Conducted at one period of time or at one time
- Selection of the subjects - done without referring or knowing the initial disease or factors ; then examined
- Disease exposure and status are concurrently evaluated among individuals in a population
- Provide snapshot view / information regarding the prevalence of a disease


POPULATION
DISEASE + , FACTOR +
DISEASE +, FACTOR -
DISEASE -, FACTOR +
DISEASE -, FACTOR -
Onset of Study

The Structure of a Cross-Sectional Study



The Structure of a Cross-Sectional Study, Continued

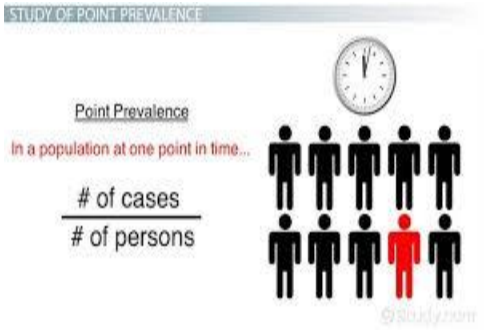


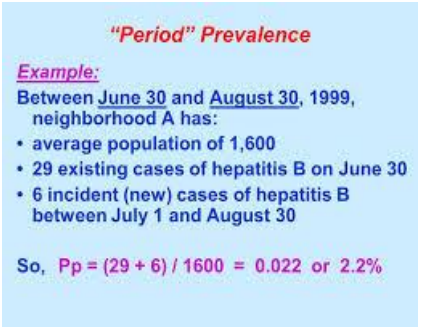
Characteristics of Cross Sectional Study

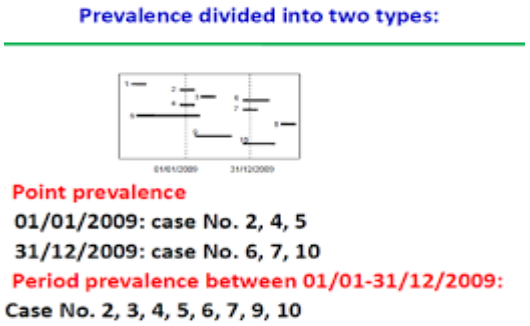
Subject	Assessed individually
Characteristics	“point-in-time” picture of health status/ health-related behaviour
Orientation	Present
Time factor	Quick
Cost	Relatively inexpensive
Follow-up period	None
Purpose	Generate hypothesis
Cause-effect relationship	Cannot be interpreted
Measure of diseases frequency and risk	Prevalence, Association
Potential problems	Can’t assess seasonal variation

Types of cross sectional study

- Descriptive cross-sectional study
  - Characterize prevalence of diseases in a specified population
  - Assessment by
    - Point prevalence
    - Period prevalence
- Analytical cross-sectional study
  - To compare the diseases differences between exposed and non-exposed
  - Examines the relationship between diseases and other variables of interest that exist in the population at the time of study

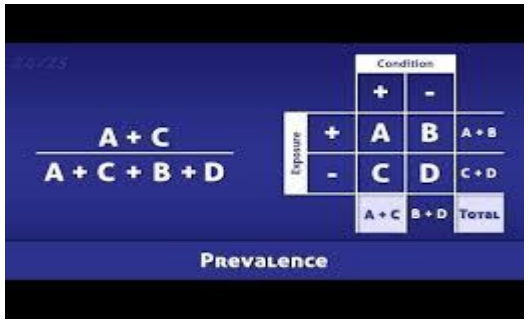


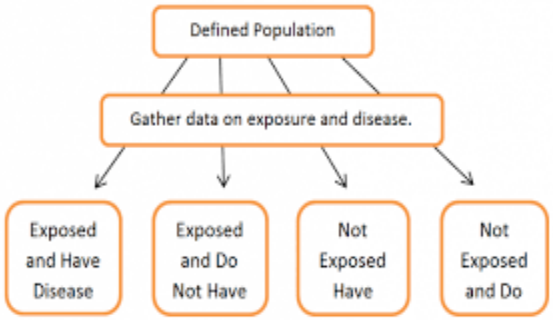




The use of cross-sectional studies

- **Prevalence survey:**
  - The studies are commonly used to describe the burden of disease in the community and its distribution.
- **Describe population characteristics:**
  - also commonly used to describe population characteristics, often in terms of person (who?) and place (where?)
  - Eg: The British National Diet and Nutrition Survey
  - To describe various age groups in the population in terms of food and nutrient intake and range of other personal and lifestyle characteristics.





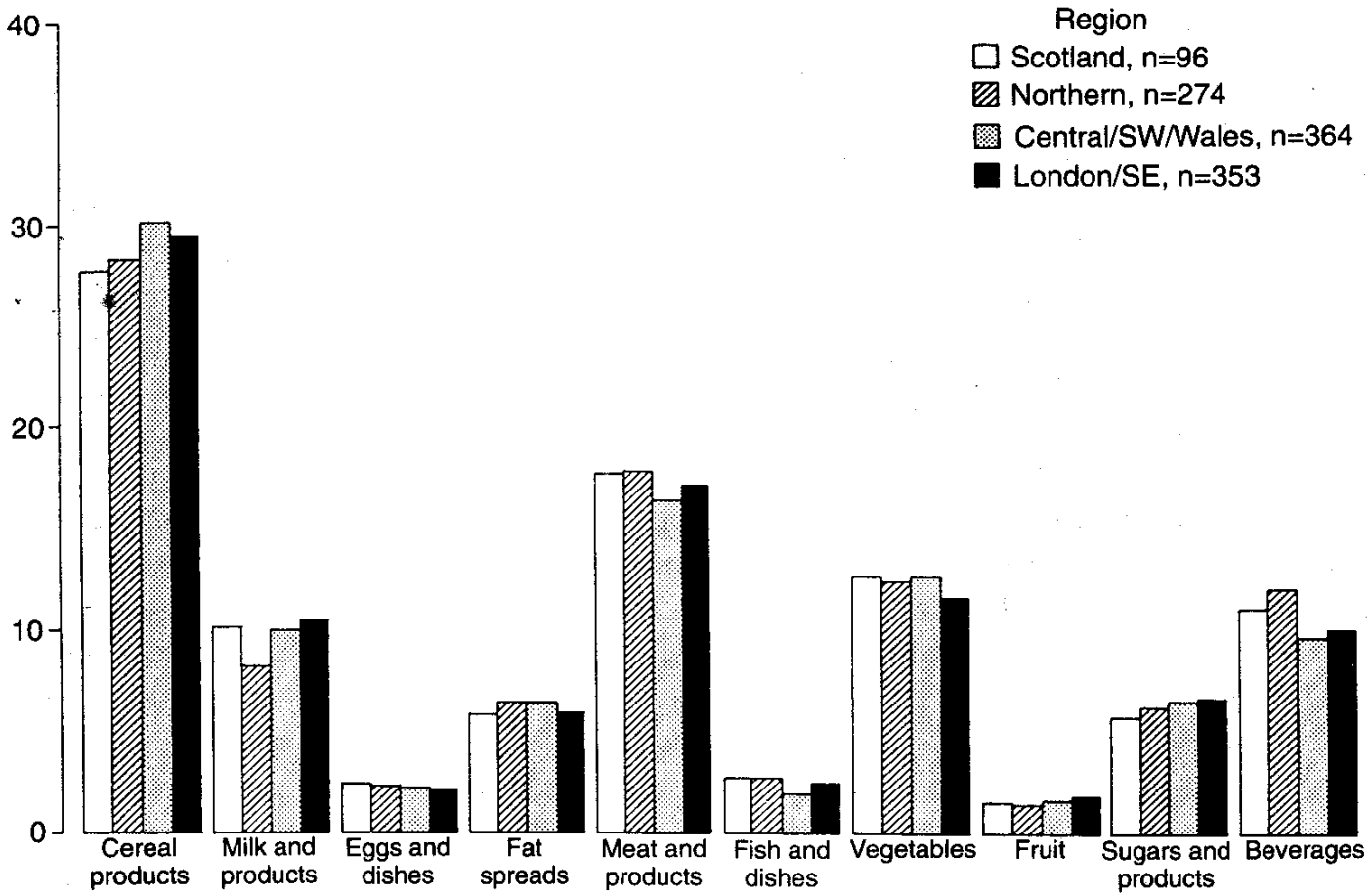


Fig. 13.2 Men: percentage of energy from main food groups by region. (Source: Gregory *et al.*[6])

- **Migrant study** :
  - Some migrant studies may full into the classification of cross-sectional studies.
  - These studies give clues as to association between genetic background and environmental exposures on the risk of disease.
  - Eg: A study of the prevalence (percentage) of coronary heart disease among men of Japanese ancestry living in Japan, Honolulu and the San Francisco Bay area
  - It showed the highest rates among those who had migrated to the United States.

- **KAP (knowledge, attitude, and practice) study:**
  - KAP studies are purely descriptive and help to build up a better understanding of the behavior of the population, without necessarily relating this to any disease or health outcome.
- **Management tool**:
  - Health service managers and planners may make use of cross-sectional survey to assess utilization and effectiveness of service.
- **Development of hypothesis**:
  - Hypotheses on the causes of disease may be developed using data from cross-sectional study survey.

ADVANTAGES

- Interview/examination needed only once: quickly done and easy
- The association obtained can produce hypothesis
- Scientific sample collection can represent the whole population
- Prevalence rate can be obtained and used to plan health needs, to know disease burden etc.
- Normally does not need high expenditure to do it

DISADVANTAGES

- Difficult to determine whether exposure comes before the disease (because exposure is evaluated at the same time) – not able to conclude the temporal sequence between exposure and disease
- Identify high proportion of prevalent cases of long duration. People who die soon after diagnosis or who recover quickly are less likely to be identified as diseased
- Poor estimates of infrequent or rare disease

Bias in Cross Sectional Study

- Measurement bias
  - Using incorrect method
- Sampling bias
  - Occur any time when the sample is not a random sample, and allow certain individual to be selected in the studies.
- Selection bias
  - May occurred if the method to get the sample not randomized
  - When the study population not represent the actual population.

    → Distortion in estimating the association between risk factor and diseases.
- Recall bias
  - When questionnaire and interview based method used in data collection.

Bias in Cross Sectional Study

- Late look bias
  - fewer pt with severe diseases detected in the study because they died before detection (Neyman bias)
- Length bias
  - People with less aggressive diseases are selected because they live longer and included in the study (old cases-longer duration of diseases)
- Lead-time bias
  - Pt with diseases was find in earlier stage by screening asymptomatic pts regardless of aggressiveness of the diseases. This will make the pts seems to live longer with the diseases.

What do you measure?

- Prevalence – point or period
- Prevalence difference
  - PD = P1 – P2
- Prevalence ratio
  - PR = P1 / P2
- Prevalence OR
  - $POR = \frac{P1/(1 - P1)}{P2/(1 - P2)}$

	Disease	No Disease	Total
Exposed	a	b	a+b
Non exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

P1= a/a+b (Prevalence of disease among the exposed

P2= c/c+d (Prevalence of disease among the non-exposed

Example

	Disease	No Disease	Total
Exposed	15	237	252
Non exposed	5	175	180
Total	20	412	432

Prevalence = 20 / 432 = 4.63%

P1=(15/252)= 0.06 P2=5/180=0.028

PD = P1 – P2 = 0.06 – 0.028 = 0.032 = 3.2%

PR = P1 / P2 = 2.14 POR = 2.07

azmawati@ppukm.ukm.edu.my

THANK YOU