Overview of study designs

Dr Munir Abu-Helalah
MD,MPH,PHD

Associate Professor of Epidemiology and Preventive Medicine
Consultant Clinical Preventive Medicine

Key references

- Bowling A. Research methods in health. Second edition. Berkshire. Open University Press. 2002
- Crombie I. Research in Health Care. Chichester. John Wiley & Sons. 1997.

Lilienfield D, Foundations of epidemiology. Third Edition. New York. Oxford University Press. 1994.

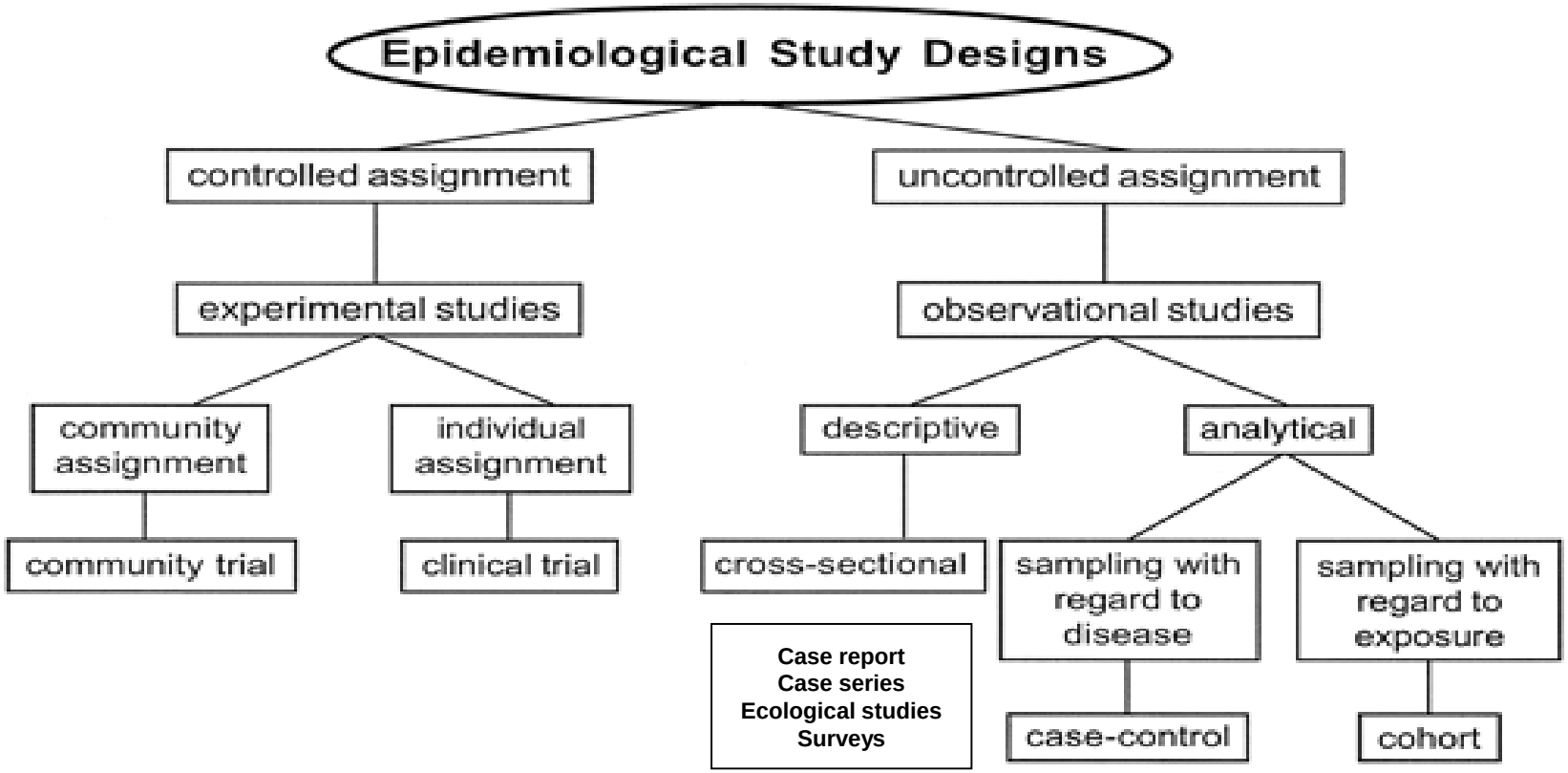
- Bhopal R. Concepts of epidemiology. New York. Oxford University Press. 2004.
- Wald N. The Epidemiological approach. Fourth Edition. London. Royal Society of Medicine Press.

Part 1

Descriptive studies

Study design: Definition

A study design is a specific plan or protocol for conducting the study, which allows the investigator to translate the conceptual hypothesis into an operational one.



Source: Waning B, Montagne M: *Pharmacoepidemiology: Principles and Practice*: http://www.accesspharmacy.com

Observational epidemiology

- Provides information about disease patterns or drug use problems by various characteristics of person, place, and time.

- It also is used by epidemiologists to generate hypotheses regarding the causes of disease or drug use problems.

Observational epidemiology

a. Descriptive

Case reports and case series

Descriptive analysis (Person place time)

Ecological (correlational)

Cross-sectional

b. Analytical

Case Control

Cohort

Epidemiological studies

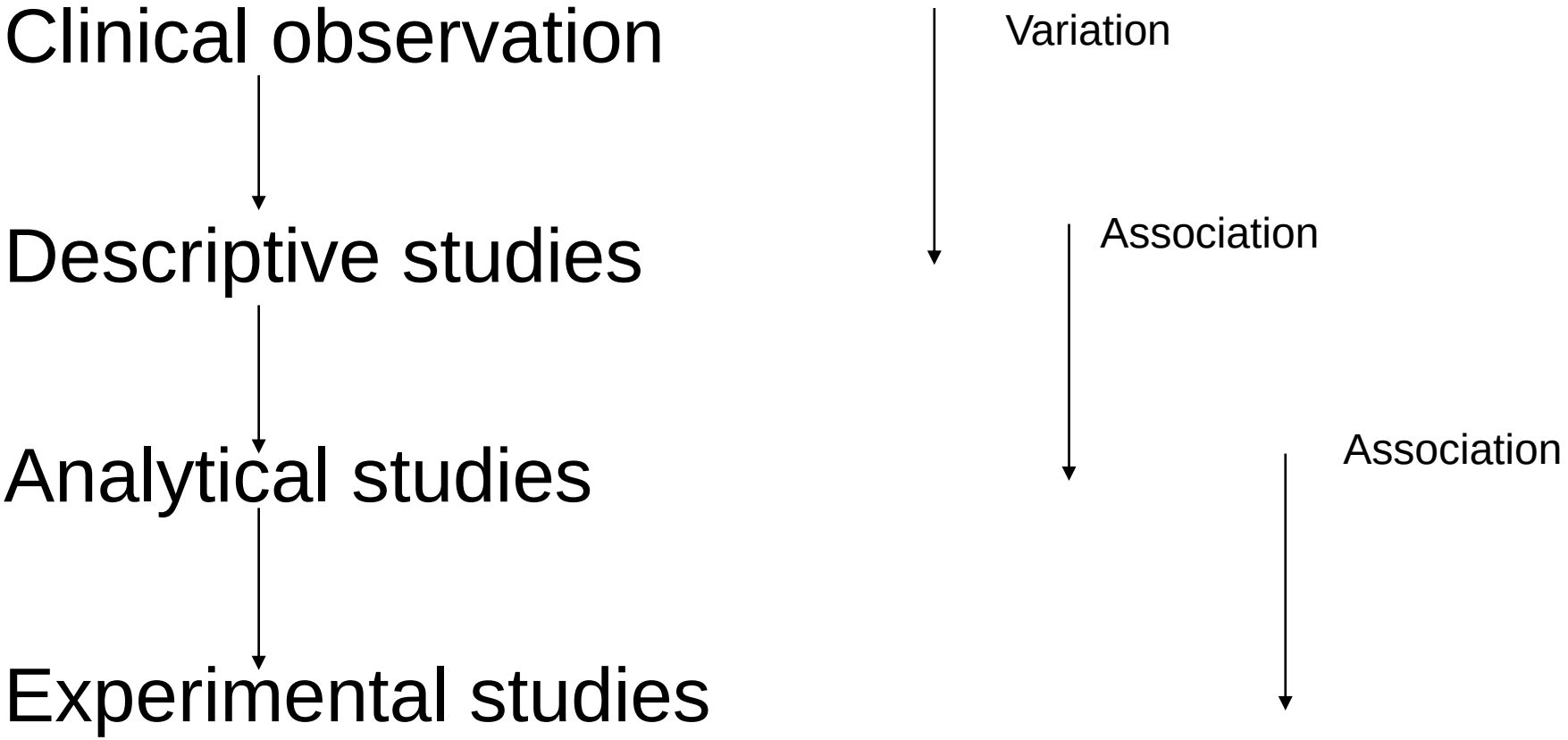
- Descriptive studies attempt to uncover and portray the occurrence of the condition or problem, whereas analytical studies determine the causes of the condition or problem.
- Investigators in observational studies may plan and identify variables to be measured, but human intervention is not a part of the process.
- Experimental studies, in contrast, involve intervention in ongoing processes to study any resulting change or difference. Epidemiological studies are also descriptive or analytical in nature.

Observational epidemiology

- Descriptive studies: provide insight, data, and information about the course or patterns of disease or drug use problems in a population or group.

- Analytical studies are used to test cause–effect relationships, and they usually rely on the generation of new data.

Epidemiological studies



Does coffee causes pancreatic cancer

I am beginning to suspect that there is an association between coffee drinking and pancreatic cancer ……

Case series

Descriptive analysis

Ecological study

Cross-sectional analysis

How to investigate this further?

Data Collection Methods

- **Primary:** where the investigator is the first to collect the data.

Sources include: medical examinations, interviews, observations, etc.

Advantage: less measurement error, suits objectives of the study better.

Disadvantage: costly, may not be feasible.

- **Secondary**: where the data is collected by OTHERS, for other purposes that those of the current study.
- Sources include: individual records (medical / employment); group records (census data, vital statistics)

Prospective vs. retrospective studies

Prospective studies

- Watches for outcomes, such as the development of a disease, during the study period and relates this to other factors such as suspected risk or protection factor(s). The study usually involves taking a cohort of subjects and watching them over a long period.
- The outcome of interest should be common; otherwise, the number of outcomes observed will be too small to be statistically meaningful (indistinguishable from those that may have arisen by chance).
- All efforts should be made to avoid sources of bias such as the loss of individuals to follow up during the study.
- Prospective studies usually have fewer potential sources of bias and confounding than retrospective studies.

Retrospective studies

- **A looks backwards and examines exposures to suspected risk or protection factors in relation to an outcome that is established at the start of the study.**
- **Many valuable case-control studies, such as Lane and Claypon's 1926 investigation of risk factors for breast cancer, were retrospective investigations.**
- **Most sources of error due to confounding and bias are more common in retrospective studies than in prospective studies.**

Comparison of Retrospective and Prospective Approaches	
Retrospective	**Prospective**
Inexpensive to conduct	Expensive to conduct
Completed in a shorter time period	Completed over a longer time period
Easier to access a larger number of subjects	More difficult to access subjects and usually requires a larger number of subjects
Allows results to be obtained more quickly	Exposure status and diagnostic methods for disease may change
Useful for studying exposures that no longer occur	Loss of subjects from the study over time may be substantial
Information and data may be less complete and inaccurate	Information and data may be more complete and accurate
Subjects may not remember past information	Direct access to study subjects enhances reliability of data

Case Reports and Case Series

Case report is detailed report by one or more clinicians of the profile of a single patient.

Example: 1961; pulmonary embolism 5 weeks after use on oral contraceptive.

Question: Are women who develop pulmonary embolism more likely to have used oral contraceptives than women who did not develop the disease?

Case Series describes the characteristics of a number of patients with a given disease.

Application: Routine surveillance activities (accumulated case reports). Striking clustering of cases may suggest emergence of new diseases or epidemics

Case report and case series

- **Clinician finds unusual features of a disease or effects of a drug, or the patient's medical history, that lead to the formulation of a new research question or hypothesis**

Hammade et al.
Journal of Medical Case Reports (2022) 16:386
https://doi.org/10.1186/s13256-022-03630-1

Journal of
Medical Case Reports

CASE REPORT Open Access

Isolated giant renal hydatid cyst with a simple renal cyst appearance: a case report

Mohammed Hammade[1*], Sami Alhoulaiby[1] and Adnan Ahmed[2]

Abstract

Background: Isolated renal hydatid cysts of the kidney are a rare occurrence that account for about 2–3% of all hydatidoses. They can stay asymptomatic for years and could have a variable presentation on imaging techniques, which results in a challenging diagnostic process.

Case presentation: We report a 22-year-old Caucasian male with a large cyst on the upper pole of the left kidney that had no septations nor membrane calcifications on computed tomography, which led to mistakenly considering it a simple renal cyst. The true diagnosis was identified intraoperatively and proven postoperatively by pathology.

Conclusions: This case highlights the importance of keeping echinococcosis in mind when treating suspected renal cysts and tumors to avoid incorrect treatment and possible content spillage, anaphylaxis, and peritoneal dissemination.

Keywords: Isolated renal hydatid cyst, Renal echinococcosis

Case reports

- The most common type of study published in the medical literature. They note unusual medical occurrences, identify new diseases, and describe adverse effects from drug therapies.

- Clinical investigators can use challenge–rechallenge data to help establish causality.
- In this approach, administration of a drug (the challenge) might be suspected of producing a specific symptom (side effect or adverse reaction).
- Administration of the drug can be stopped to observe whether the side effect or adverse reaction diminishes.
- If it does, then administration of the drug can be resumed (the rechallenge) to observe whether the effect returns, suggesting a possible relationship between the two events.

Case-series:
Clinical case series

- Usually a coherent and consecutive set of cases of a disease (or similar problem) which derive from either the practice of one or more health care professionals or a defined health care setting, e.g. a hospital or family practice.

Case-series: Clinical case series

- A case-series is, effectively, a register of cases.
- Analyse cases together to learn about the disease.
- Clinical case-series are of value in epidemiology for:
  - Studying symptoms and signs
  - Creating case definitions
  - Clinical education, audit and research

Case series:
Natural history and spectrum

- **Helps professionals can build up a picture of the natural history of a disease**

Case series:
Natural history and spectrum

- **Population case-series is a systematic extension of this series but which includes additional cases, e.g. those dying without being seen by the clinicians.**

- **Add breadth to the understanding of the spectrum and natural history of disease.**

Case series: Limitations

Usually we cannot estimate the prevalence or incidence rate

- **Breast cancer registry in Jordan: We cannot provide incidence or prevalence rates without:**

1. **Population size**
2. **Time- period of data collection**
3. **All cases of breast cancer are registered**

<u>No control group for comparison</u>

Case series: Population

- Case-series can provide the key to sound case control and cohort studies and trials
- Design of a case-series is conceptually simple
- Defines a disease or health problem to be studied and sets up a system for capturing data on the health status and related factors in consecutive cases

Congenital Rubella Syndrome: The classic description of a series of infants born with congenital cataracts, some with additional cardiac abnormalities, in Australia in 1941. This led Gregg in Sydney to postulate a causal link between a severe epidemic of rubella that had occurred six to nine months before the children were born and the subsequent abnormalities. It is now well known that if a woman develops rubella during pregnancy it may affect her unborn baby.

Tuberculosis is a disease relatively frequent in renal transplant patients, presenting a wide variety of clinical manifestations, often involving various organs and potentially fatal. Gastrointestinal tuberculosis, although rare in the general population, is about 50 times more frequent in renal transplant patients. Intestinal tuberculosis has a very difficult investigational approach, requiring a high clinical suspicion for its diagnosis. Therapeutic options may be a problem in the context of an immunosuppressed patient, requiring adjustment of maintenance therapy. The authors report two cases of isolated gastro-intestinal tuberculosis in renal transplant recipients that illustrates the difficulty of making this diagnosis and a brief review of the literature on its clinical presentation, diagnosis, and therapeutic approach.

Dehneh et al. Journal of Medical Case Reports (2022) 16:371
https://doi.org/10.1186/s13256-022-03609-y

Journal of
Medical Case Reports

CASE REPORT Open Access

Syrian females with congenital adrenal hyperplasia: a case series

Nada Dehneh[1*], Rami Jarjour[2,3], Sahar Idelbi[4], Assad Alibrahem[4,5] and Sahar Al Fahoum[1]

Abstract

Background: One of the most common types of congenital adrenal hyperplasia is an autosomal recessive disorder with 21-hydroxylase deficiency. The classical form, defined by cortisol insufficiency, is accompanied by prenatal androgen excess causing variable masculinization degrees of external genitalia in babies with a 46, XX karyotype.

Cases presentation: These five case reports highlight the management of Syrian females aged between 0 and 32 years with congenital adrenal hyperplasia. Two of the patients have been raised as males, while two had reconstructive surgery and one had hormonal therapy. Becoming mother was achieved by two patients

Conclusion: The integrated treatment of females with classical congenital adrenal hyperplasia CAH, which includes appropriate surgical procedures and controlled hormonal therapy, gives these females the opportunity to live as they are, and perhaps as mothers in the future.

Keywords: Congenital adrenal hyperplasia, Syria, Case report

Ecological studies

Are studies in which information on the characteristics and/or exposures of individual members of the population groups are generally not obtained. Existing statistics are used to compare the mortality or morbidity experience of one or more populations with some overall index exposure. care is needed to avoid the 'ecological fallacy' where inappropriate conclusions are made from ecologic data

Ecological studies

- These studies are used to describe disease or drug use problems in relation to some factor of interest.

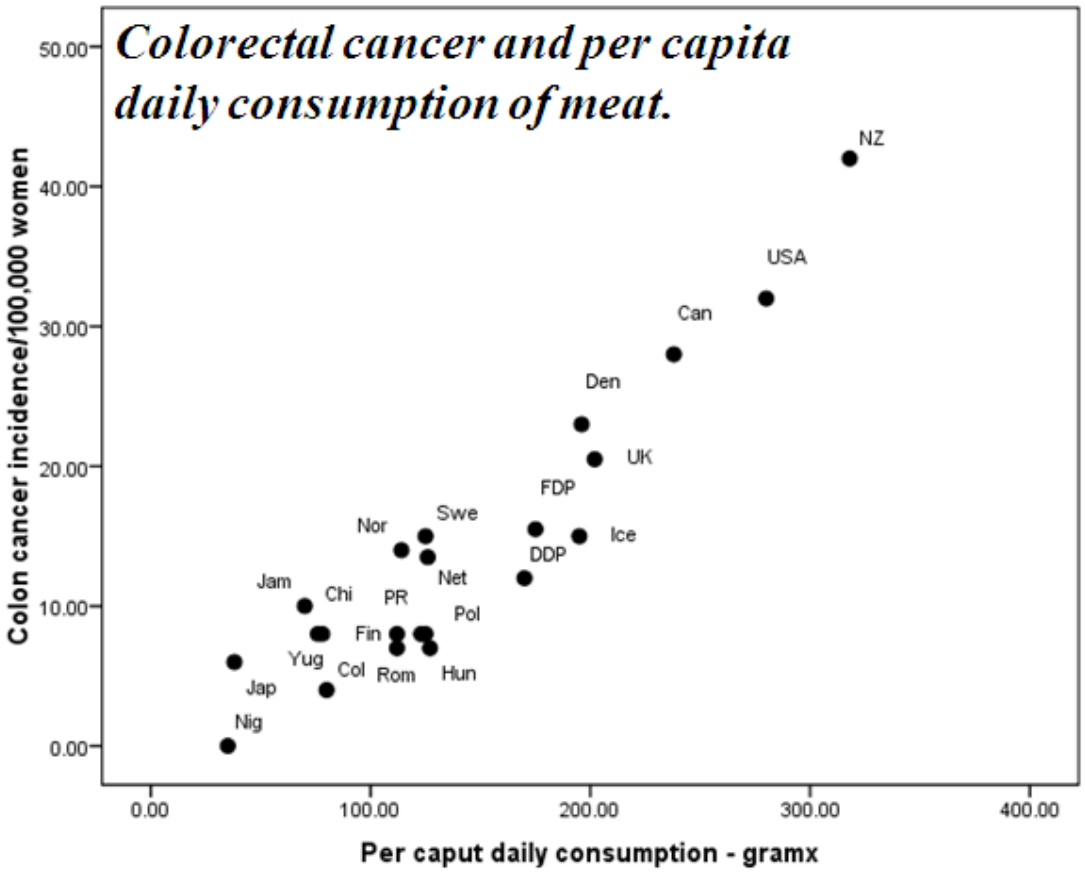
Comparing cigarette consumption with rates of cancer

Comparing Alcohol consumption with coronary heart disease mortality

- Ecological studies are the first identified strong relationships between disease and behavior.

Ecological studies

- In ecological studies the unit of analysis is some aggregate individuals rather than individual persons
- Geographic areas or time period are often used as a basis for defining aggregates
- The analysis centers on determining whether the ecological units with a high frequency of exposure are also unit with a high frequency of disease (+ve correlation) or a low frequency of disease (- ive correlation)



Adapted from: *Int. J. Cancer 15:617, 1973*

Ecological (correlational studies)

- look for associations between exposures and outcomes in populations rather than in individuals.
- They use data that has already been collected.

- The measure of association between exposure and outcome is the correlation coefficent *r*.
- This is a measure of how linear the relationship is between the exposure and outcome variables. (Note that correational is a specific form of association and requires two continuous variables)

Ecological (correlational studies)

Advantages of an ecological study

1. An ecological study is quick and cheap to conduct.
2. It can generate new hypotheses.
3. It can identify new risk factors.

Ecological (Correlational studies)

Disadvantages:

1. It is unable to control for confounding factors. This is often referred to as 'ecological fallacy', where two variables seem to be correlated but their relationship is in fact affected by cofounding factor(s).
2. It cannot link exposure with disease in individuals as those with disease may not be expose.
3. Its use of average exposure levels masks more complicated relationships with disease.
4. Its units of study are populations not individuals. Therefore, the disease rates linked with population characteristics and the association observed at group level does not reflect association at individual level.

Ecological (correlational studies)

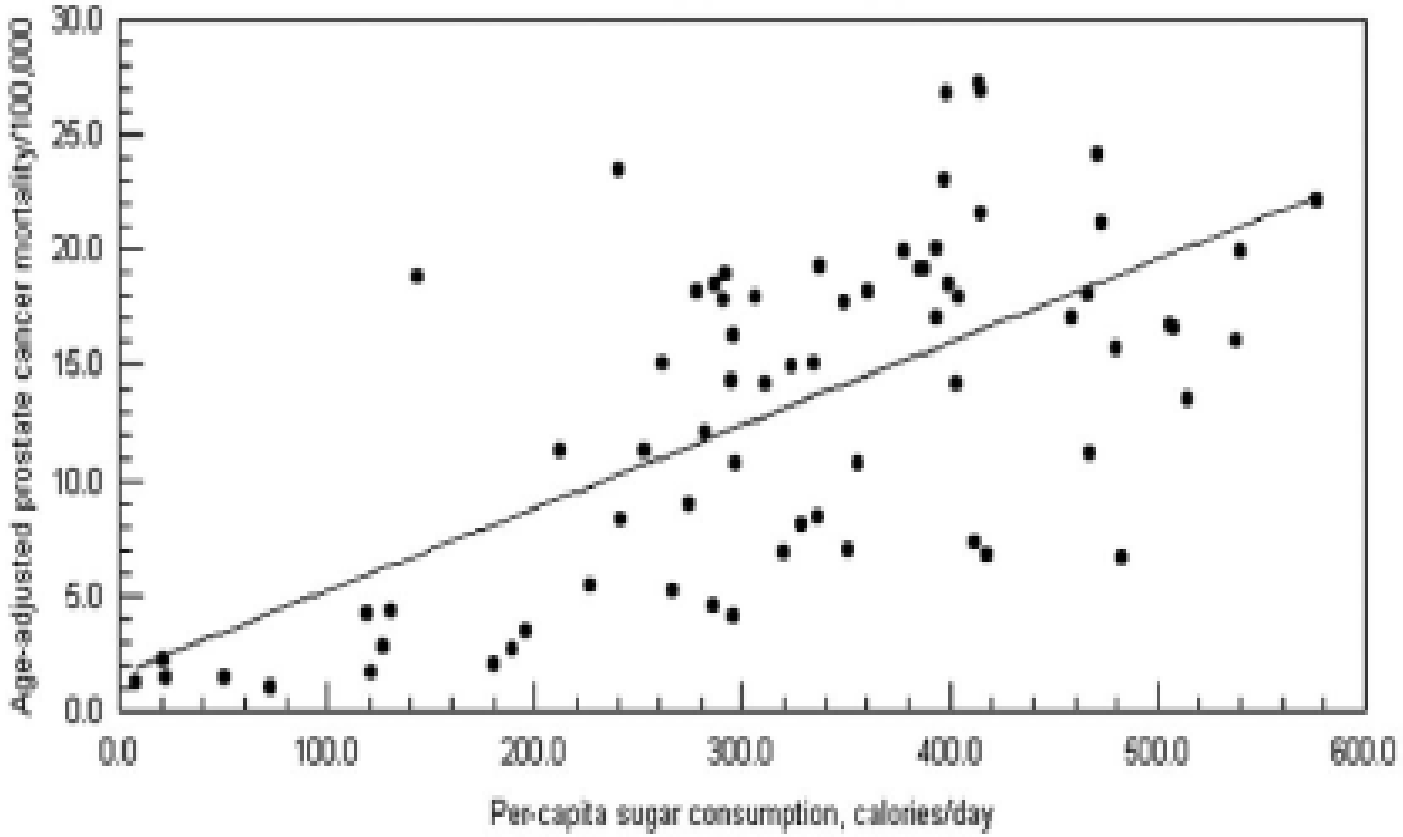


Fig. 1. Prostate cancer mortality versus sugar consumption in 71 countries.

Descriptive epidemiology

- There are many problems with descriptive methods.
- In case reports and case series, there is no control group.
- For correlation studies: there are confounding factors that might mask the true impact of risk factors.
- Correlation studies present only a snapshot of the problem, such as disease or drug use, in a population.

CROSS-SECTIONAL STUDY DESIGN

- **Sometimes called *prevalence studies*.**
- **They are studies of total populations or population groups in which information is collected about the present and past characteristics, behaviors, or experiences of individuals.**
- **There are a number of advantages in performing a cross-sectional study. These studies involve a single data collection and, thus, are less expensive and more expedient to conduct.**

CROSS-SECTIONAL STUDY DESIGN

- **Emphasis is on differences between groups at one point in time.**
- **They provide a one-time glimpse at the study population, showing the relative distribution of conditions, diseases, and injuries—and their attributes—in a group or population.**

Cross-sectional (or prevalence) studies

Are studies in which a defined population is surveyed and their disease or exposure status determined at one point in time

- **The prevalence rates of disease in the whole population as well as in those with and without the exposure under investigation can be determined**

- **Cross-sectional studies are generally not suitable for a disease which is rare or of short duration as few people will have the disease at any one point in time**

Cross-sectional studies

- More effective in identifying chronic diseases and problems
- Less effective in identifying communicable diseases of short incubation periods and short durations.

Cross-sectional (or prevalence) studies

- **It is often difficult to separate cause and effect as the measurement of exposure and disease at any one point in time**

- **Because of this limitation, cross-sectional studies are useful when investigating exposures which do not change e.g genetic characteristics such as ABO blood group and HLA**

- **Cross-sectional studies are often used as an initial exploration of a hypothesis prior to conducting a case-control or follow-up study**

CROSS-SECTIONAL STUDY DESIGN

- They provide information and data useful for the planning of health services and medical programs.

- They are based on a sample of the whole population and do not rely on individuals presenting themselves for medical treatment

CROSS-SECTIONAL STUDY DESIGN

- Sample size:

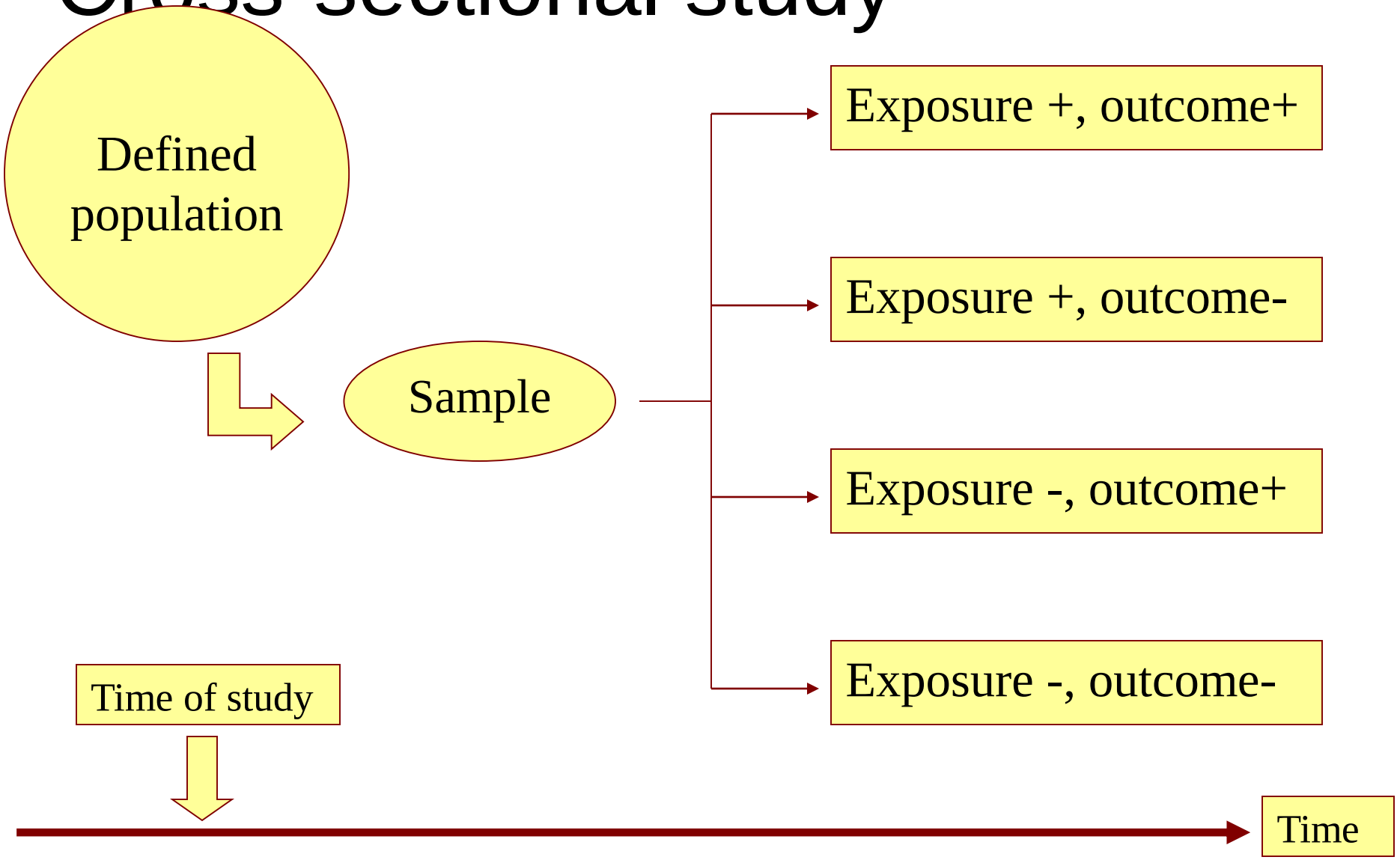
1. Question or primary & secondary outcomes
2. Population size
3. Prevalence of condition of interest in the population
4. Distribution of the condition (for example hypothyroidism is common among women age 50 to 70 but less common amongst men at this age group).

   Therefore we need a large sample from men in the general population to get men with hypothyroidism. In this case we stratify for gender.

Cross-sectional study

- Exposure and outcome are assessed simultaneously among individuals in a defined population, thus at one point in time
- No sampling of individuals based on a exposure or an outcome

Cross-sectional study



Two by two table

Exposure	**Outcome** Yes	No	Total
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b + c + d

Prevalence of outcome in exposed = a / a + b

Prevalence of outcome in non-exposed = c / c + d

Prevalence Rate Ratio (PRR) = $= \frac{a / a + b}{c / c + d}$

Cross-sectional study

Prevalence of and Factors Associated With Persistent Pain Following Breast Cancer Surgery

JAMA. 2009;302(18):1985-1992

Objective To examine prevalence of and factors associated with persistent pain after surgical treatment for breast cancer.

Design, Setting, and Patients A nationwide cross-sectional questionnaire study of 3754 women aged 18 to 70 years who received surgery and adjuvant therapy (if indicated) for primary breast cancer in Denmark between January 1, 2005, and December 31, 2006. A study questionnaire was sent to the women between January and April 2008.

Cross-sectional study

Chemotherapy	Outcome		Total
	With pain	Without pain	
Yes	664	556	1220
No	879	1088	1967
Total	1543	1644	3187

Prevalence of pain among chemotherapy = 664/ 1220 = 54.4%

Prevalence of pain among no chemotherapy = 879 / 1967 = 44.7%

Prevalence Rate Ratio (PRR) = = 54.4 / 44.7 = 1.22

Cross-sectional survey of CHD among male by physical activity

	Number examined	Number with CHD	prevalence
Not physically active	**89**	**14**	**157.2/1000**
Physically active	**90**	**3**	**33.3/1000**

From: *BRCA1* and *BRCA2* genes mutations among 200 high risk breast cancer patients in Jordan

Category	Number of patients	Prevalence (total 200)
Recurrent mutations		
BRCA1 Positive	15	7.50%
BRCA2 Positive	14	7.00%
BRCA1 or BRCA2 Positive	29	14.50%
Possible (recurrent and novel) mutations		
BRCA1 Positive	7	3.50%
BRCA2 Positive	14	7.00%
BRCA1 or BRCA2 Positive	21	10.50%
Recurrent and novel (VUS and pathogenic) mutations		
BRCA1 Positive	15	7.50%
BRCA2 Positive	21	10.50%
BRCA1 or BRCA2 Positive	36	18.00%

Abu-Helalah et al. https://www.nature.com/articles/s41598-020-74250-2

Cross-sectional studies

- Seasonal variations of disease are not well represented in cross-sectional studies except if the duration of the study allows such comparison
- In the example below, studying RTA in October would not provide a valid result for incidence of RTA in whole year and does not allow identifying seasonal variations in the RTA
- Road traffic accidents by month of accident, Slovenia, average 2003-2006

Cross-sectional studies: advantages

- **Quick**
- **Many associations can be studies**
- **Data on all variables is only collected once.**
- **Sample size depends on the question**
- **Standard measures used**
- **Prevalence estimated**
- **The prevalence of disease or other health related characteristics are important in public health for assessing the burden of disease in a specified population and in planning and allocating health resources.**
- **Good for descriptive analyses and for generating hypotheses**

Cross-sectional studies

Disadvantages:

- They cannot show cause–effect relationships.

Difficult to determine whether the outcome followed exposure in time or exposure resulted from the outcome.

- If the sample is not representative, results are representative only of the individuals who participate in the study

Example prevalence of sickle cell anaemia in the Easter region of the KSA does not represent the who country.

- This design is not effective if the level of disease rate is very small.
- Not suitable for studying rare diseases or diseases with a short duration.
- Unable to measure incidence unless the duration of study allows.
- Associations identified may be difficult to interpret.
- Susceptible to bias due to low response and misclassification due to recall bias.

Study Design: Part 2

Dr Munir Abu-Helalah
MD,MPH,PHD
Associate Professor of Epidemiology and Preventive Medicine

Case-control studies

Are studies in which a group of people with a particular disease (the cases) are compared with a group of people without the disease (the controls). The purpose of the comparison is to determine whether, in the past, the cases have been exposed more (or less) often to a specific factor than the controls

- This type of study is done to identify factors that could be responsible for the development of a disease or drug use problem.

CASE-CONTROL STUDIES

- The direction of time

- Cases identified now
- Data on past events collected

CASE-CONTROL STUDY DESIGN

- Designed to assess association between disease occurrence and exposures (e.g., causative agents, risk factors) suspected of causing or preventing the disease.

Case-control studies

- A group of people with a disease are compared to a group without the disease from the same population.
- Compare exposure to risk factors in both groups
- Able to look at many different possible risk factors
- Able to study diseases with a long latency period
- Most common analytic study design seen in the medical literature today

Case-control studies

- In general, the cases included in a case-control study include people with one specific disease only
- But, a case-control study can provide information on a wide range of possible exposures that could be associated with that particular disease
- Useful for the study of rare diseases
- Not suitable for the study of rare exposure
- Relatively small and inexpensive
- Takes a relatively short time to complete
- Can test current hypotheses
- Cannot measure disease incidence

CASE-CONTROL STUDIES

- Cases have the disease of interest

Eg. Cerebral palsy

- Controls do not have the disease

Eg. Healthy babies born at the same time

Two by two table

Exposure	**Outcome** Yes	No	Total
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b + c + d

Odds of outcome in exposed = a / b

Odds of outcome in non- exposed = c / d

Outcome odds ratio = (a/b) / (c/d) = ad / bc

Case-control study: challenges

- Selecting cases
  - Eligibility
- Selecting controls
  - Representativeness
- Exposure assessment
  - Accurate

Case-control study

Early life exposure to diagnostic radiation and ultrasound scans and risk of childhood cancer: case-control study

BMJ 2011;342:d472

Objective To examine childhood cancer risks associated with exposure to diagnostic radiation and ultrasound scans in utero and in early infancy (age 0-100 days).
Design Case-control study.
Setting England and Wales.
Participants 2690 childhood cancer cases and 4858 age, sex, and region matched controls from the United Kingdom Childhood Cancer Study (UKCCS), born 1976-96.
Main outcome measures Risk of all childhood cancer, leukaemia, lymphoma, and central nervous system tumours, measured by odds ratios.

Case-control study: example

Radiation	**Case**	**Control**	Total
Yes	120	185	305
No	2570	4672	7242
Total	2690	4857	7547

Odds of outcome in exposed = 120 / 185 = 0.65

Odds of outcome in non-exposed = 2570 / 4672 = 0.55

Outcome odds ratio = (a/b) / (c/d) = 1.18

CASE-CONTROL STUDIES

Methods of data collection

Case-note review: Completeness

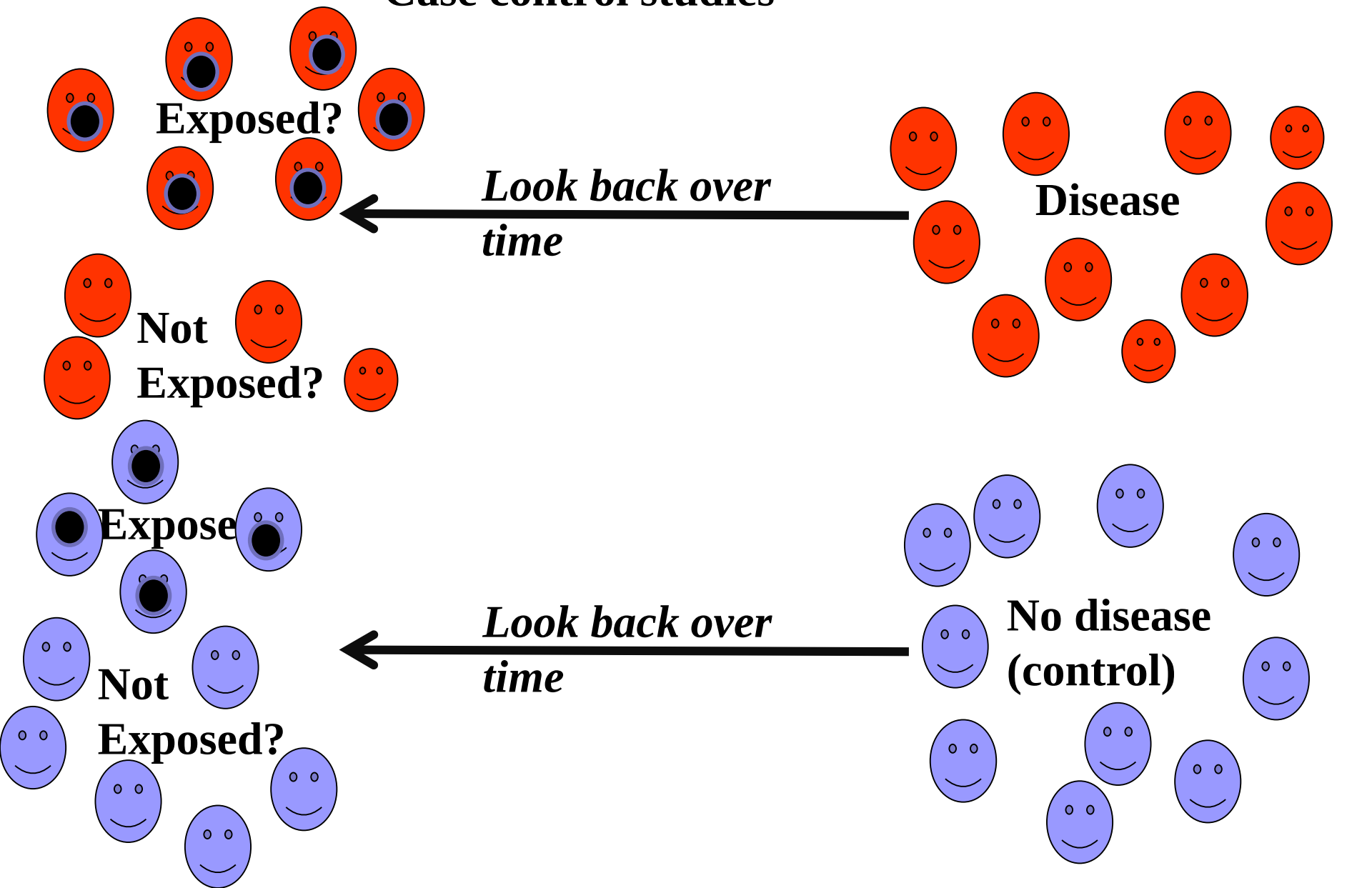
Postal questionnaire: response rate

Interview: Detailed information

CASE-CONTROL STUDY DESIGN

- More efficient than a cohort study because a smaller sample size is required.
- One key feature of a case-control study, which distinguishes it from a cohort study, is the selection of subjects based on disease status.
- Controls are chosen from the same population yielding the cases

Case control studies



CASE-CONTROL STUDIES

Strengths

- Suited to study disease with long latency periods, but can be used in outbreaks investigations
- Optimal for rare diseases
- Efficient in terms of time and costs: relatively quick and inexpensive
- Allows for evaluation of a wide range of possible causative factors that might relate to the disease being studied
- Odds ratio estimated

CASE-CONTROL STUDIES

Limitations

- Very susceptible to bias (especially selection and recall bias) as both the disease and the exposure have already occurred when participants enter the study. Cases and controls might not be representative of the whole population
- We cannot calculate incidence or prevalence rate of disease
- We cannot be certain that exposure came before disease
- Choice of controls difficult
- Controls do not usually represent non-exposed population
- Past records incomplete
- No absolute risk estimates

Design of case control studies

- **Comparability:Two groups must be as similar to each other as possible so selection of controls is very important. Controls must be as similar as possible to cases – except that they do not have the outcome (disease).**
- **Outcome (disease) must be very clearly defined. (Diagnostic criteria must be clear)**
- **Use objective data about exposure status wherever possible, to reduce the risk of bias**

How many controls?

- **control-to-case ratio is 1 : 1**
  is the optimal when the number of available cases and controls is large and the cost of obtaining information from both groups is comparable
- **control-to-case ratio is 1 : n**
  When the number of cases is limited or when the cost of obtaining information is greater for cases or controls

- **As the number of controls per case increases, the power of the study also increase**

- **It is not recommended that this ratio increase beyond 4 : 1**

Bias

Bias is any systematic error in an epidemiological study that results in an incorrect estimate of the association between exposure and risk of the outcome

- **Selection bias:** inappropriate controls
- **Observation bias**
  - **Subject and recall bias:** eg recall bias of mothers with cerebral palsy babies
  - **Interviewer bias: blind if possible**
  - **Misclassification**

CASE-CONTROL STUDY DESIGN

- **Selecting Cases and Controls**
- Identification and collection of cases involves specifying the criteria for defining a person as a case—in other words, as having the disease (also called *case definition*).
- This definition consists of a set of criteria, also called *eligibility criteria,* for inclusion in the study. There also are criteria for exclusion from the study.

CASE-CONTROL STUDY DESIGN

- The next step is selection of the controls.
- Controls are chosen from the source population.
- The source population is usually defined by geographic area. It is important to select controls so that participation does not depend on exposure.

CASE-CONTROL STUDY DESIGN

Source of controls

- The ideal situation is a random sample from the same source population as the cases.
- Investigators may use more than one control group.
- Controls can be selected by sampling:

  The general population in the same community; the hospital community (patients in the same hospital); individuals who reside in the same block or neighborhood; and spouses, siblings, or associates (schoolmates, co-workers) of the cases.

Obtaining cases and controls for case control studies

Study	Source of cases	Source of controls
PROM (premature rupture of membrane)	Hospital patients	Hospital patients
Rheumatoid arthritis	Outpatient clinic	Other outpatient clinic
Cervical screening	GP register	GP register

CASE-CONTROL STUDY DESIGN

Matching Cases and Controls

- **Matching is a popular approach to control for confounding and selection bias in case-control studies.**
- **Matching cases and controls helps to ensure that these groups are similar with respect to important risk factors, thereby making case-control comparisons less subject to confounding or selection bias.**

CASE-CONTROL STUDY DESIGN

Prior exposure to the risk factor(s) of interest

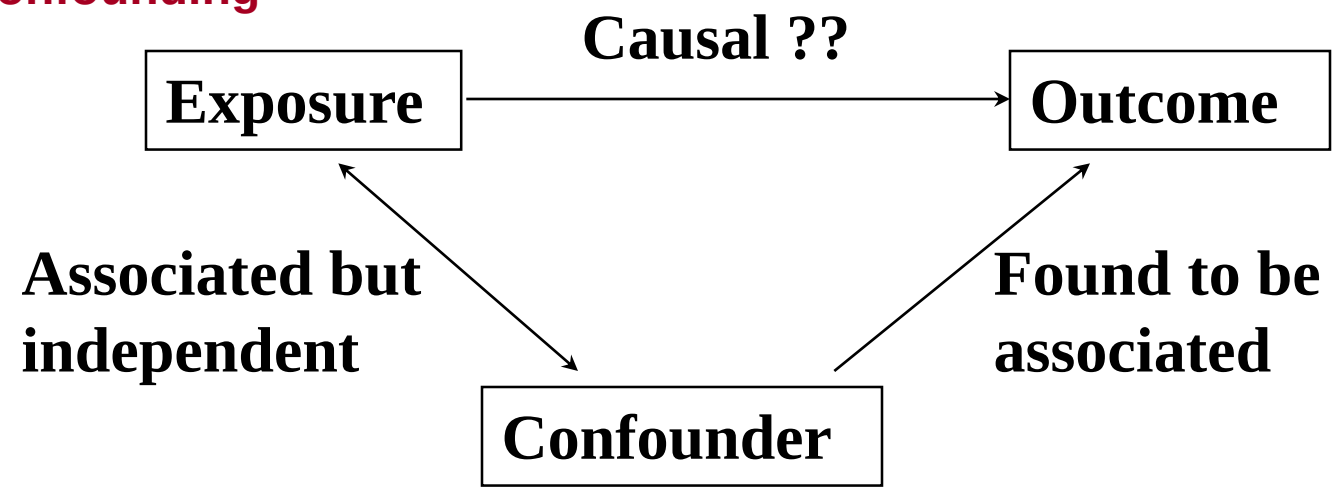
- Once cases and controls are selected, information must be collected on prior exposure to the risk factor(s) of interest.

- Interviews and questionnaires are the most common means of determining a subject's exposure history and medical records review is another source
- The most objective means for characterizing exposure is the use of a biological marker.

Confounding

A confounding factor is one that is associated with the exposure and that independently affects the risk of developing the outcome, but that is not an intermediate link in the causal chain between the exposure and the outcome under study

Matching - often used in case-control studies to decrease confounding



Confounding

Matching Cases and Controls

- For example, if age and sex are the matching variables, then a 35 year old male case is matched to a 35 year old male control
  - Pair matching (one to one individual matching)
- The use of matching usually requires special analysis techniques (e.g. matched pair analyses and conditional logistic regression)

CASE-CONTROL STUDY DESIGN

The disadvantages of matching include

(1) **It is time consuming and expensive**

(2) **Some potential cases and controls may be excluded because matches cannot be made**

(3) **Unmatched cases and controls must be discarded**

(4) **Matched variables cannot be evaluated as risk factors in the study population**

(5) **Continuous matching categories may be too broad, and residual case control differences may persist.**

CASE-CONTROL STUDY DESIGN

- **Data Analysis**
- Data collection and analysis are based on whether the case-control study involves a matched or unmatched design. The measure used typically in case-control studies is the odds ratio.
- **Odds ratio (OR):** odds of a particular exposure among people with a specific condition divided by the corresponding odds of exposure among people without the condition under study

Odds Ratio (OR)

$$OR = \frac{\text{Odds of exposure}_{\text{cases}}}{\text{Odds of exposure}_{\text{controls}}}$$

Cohort studies

Cohort (or follow-up) studies

- **Are studies in which people are identified and grouped with respect to whether or not they have been exposed to a specific factor.**

- **The groups are followed up over time to determine whether the incidence of a particular disease is any greater (or less) in the exposed group than in the non-exposed group.**

Cohort study examples:

- Life expectancy of cerebral palsy children
- Fine needle breast biopsy and breast cancer
- Aspirin intake and colorectal cancer

Cohort study:
Primary purposes

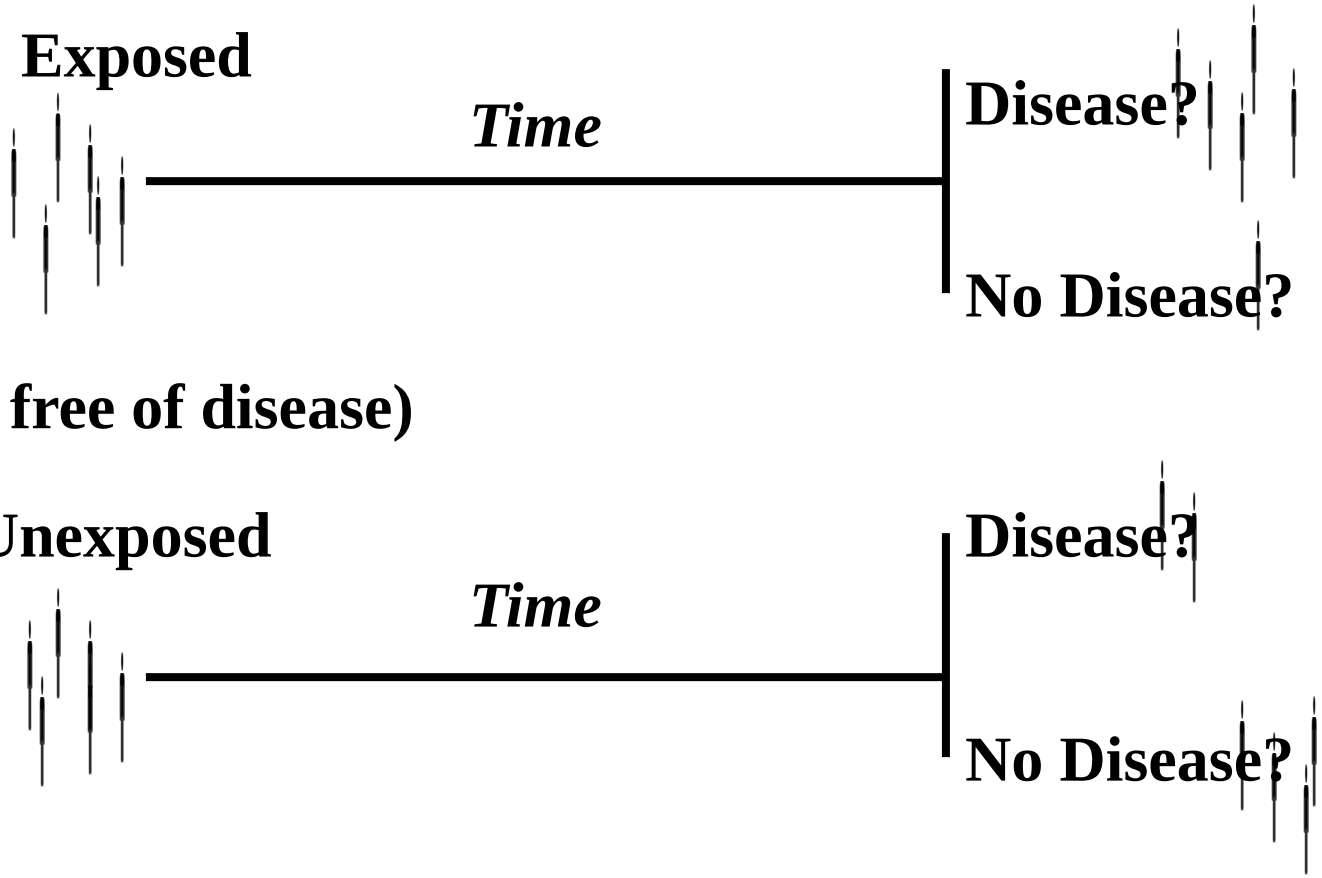
- **Descriptive (measures of frequency)**

– To describe the incidence rates of an outcome over time, or to describe the natural history of disease

- **Analytic (measures of association)**

– To analyze associations between the rates of the outcomes and risk factors or predictive factors

Cohort studies



COHORT STUDY DESIGN

- This design is the best observational one for establishing cause–effect relationships. Prevention and intervention measures can be tested and affirmed or rejected.
- Cohort studies take into account seasonal variation, fluctuations, or other changes over a longer period.
- Objective measures of exposure, such as biological markers, are preferred over subjective measures.

COHORT STUDY DESIGN
Strengths

- We can measure incidence of disease in exposed and unexposed groups
- Can get a temporal (time related) sequence between exposure and outcome as all individuals must be free of disease at the beginning of the study.
- Good for looking at effects of rare exposures.
- Allows for examination of multiple effects of a single exposure.
- Not open to bias as much as other types of study
- Direct calculation of the risk ratio or relative risk is possible.
- Provide information on multiple exposures

COHORT STUDY DESIGN

Limitations:

- Not efficient for rare diseases
- Can be expensive and time-cosuming
- Large sample
- Drop-out biases

  If study goes over many years, can get considerable loss to follow up. This can 'dilute' results or lead to bias, and therefore the validity of result can be seriously affected

- Locating subjects, developing tracking systems, and setting up examination and testing processes can be difficult.
- Changes over time in diagnostic methods, exposures, or study population may lead to biased results.

Cohort study: Example

Hypertension as a risk factor for spontaneous intracerebral hemorrhage

Physical Activity and Incident Cognitive Impairment in Elderly Persons

ARCH INTERN MED/VOL 170 (NO. 2), JAN 25, 2010

Background: Data regarding the relationship between physical activity and cognitive impairment are limited and controversial. We examined whether physical activity is associated with incident cognitive impairment during follow-up.

Methods: As part of a community-based prospective cohort study in southern Bavaria, Germany, 3903 participants older than 55 years were enrolled between 2001 and 2003 and followed up for 2 years. Physical activity (classified as no activity, moderate activity [<3 times/wk], and high activity [≥3 times/wk]), cognitive function (assessed by the 6-Item Cognitive Impairment Test), and potential confounders were evaluated. The main outcome measure was incident cognitive impairment after 2 years of follow-up.

Cohort study

Physical activity	**Cognitive impairment** Yes	No	Total
Moderate	10	990	1000
None	100	900	1000
Total	110	1880	2000

Risk of outcome in exposed (not active) = 100/1000 = 10%

Risk of outcome in non-exposed (active)=10/1000 =1%

Relative risk 10%/1%=10 =

Design of cohort studies

1. Research question must be clear
2. Set the sample size
3. Set the follow-up period (immediate, short term and long term)
4. Specify study group Sample must be representative of the population you are studying
5. All participants should be free of the outcome (disease) at the beginning of the study
6. Must be able to get correct information about exposure status easily
7. Measure the outcome
8. Comparison group must be as similar as possible to exposed group
9. Put measures in place to reduce loss to follow up if possible

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Influenced by a variety of factors including:

1. Type of exposure being investigated
2. The frequency of the exposure in the population
3. The accessibility of subjects.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Exposed and unexposed subjects must be free of the outcome of interest at the start of the study and equally susceptible to developing the outcome during the course of the study.

- If some subjects already have the outcome (e.g., disease) at the onset, then the temporal relationship between exposure and outcome becomes obscured.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Each subject must rigidly satisfy the criteria for inclusion in the cohort study, and he or she should not be excluded from subsequent analysis because of any change in exposure status during follow-up.
- The degree of surveillance should be similar in exposed and unexposed groups.
- Frequency of examination and duration of follow-up depend on the type of exposure and the outcome under investigation.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Both groups should be accessible and available for follow-up.
- Multiple comparison groups for exposed subjects chosen in different ways may reinforce the validity of findings.

Types of cohorts

- **Birth cohort : all individuals in a certain geographic area born in the same period (usually a year)**
- **Inception cohort: all individuals assembled at a given point based on some factor, e.g. where they live or work**
- **Exposure cohort: individuals assembled as a group based on some common exposure**
  - **e.g. smokers**
  - **e.g. radiation**

Healthy worker effect

phenomenon of workers usually exhibiting overall death rates lower than those of the general population due to the fact that the severely ill and disabled are ordinarily excluded from employment.

COHORT STUDY DESIGN

- Measurement of exposures should be based on intensity, duration, regularity, and variability.
- Some exposures are acute, one-time episodes never repeated in a subject's lifetime.
- Other exposures are long term, such as cigarette smoking or use of oral contraceptives.
- Exposures may also be intermittent.

COHORT STUDY DESIGN

Retrospective cohorts

- Uses information on prior exposure and disease status.
- All of the events in the study have occurred and conclusions can be drawn more rapidly.
- Costs can be lower
- May be the only feasible one for studying effects from exposures that no longer occur, such as discontinued medical treatments.
- The main disadvantage of a retrospective cohort study is that the investigator must rely on existing records or subject recall.

Ambidirectional Cohort

- Data collected both retrospectively and prospectively on the same cohort to study short and long term effect of exposure

COHORT STUDY DESIGN
Loss during follow-up

- Following subjects over a long period of time can lead to a variety of problems.
- Dropouts and losses of subjects to follow-up are major problems in cohort studies.
- Subjects may move away or leave the study for other reasons, including deaths from other causes than the disease under investigation.
- If losses to follow-up are significant during the study, then the validity of the results can be seriously affected.

COHORT STUDY DESIGN
Changes in exposure status

- It is also possible for exposure status to change during the course of the study.
- The exposure under study may be subject to variation over time.
- For example, cigarette smokers may quit, or employees may change jobs; therefore, their level of exposure to occupational hazards changes.

COHORT STUDY DESIGN
Analysis

- Collection and analysis of data on the population subgroups, based on exposure, are divided according to variables of interest, like analysis in a cross-sectional study.
- Rates for subgroups are then calculated and compared.
- Data from cohort studies are analyzed in terms of relative risk and attributable risk fractions.

COHORT STUDY DESIGN

Midpoint analysis

- Occurs when, at a defined point in time in the study, all data collected to that point are analyzed so a decision can be made to stop or continue the study.

Nested case-control study

Case-control within a cohort study

Framingham Heart Study

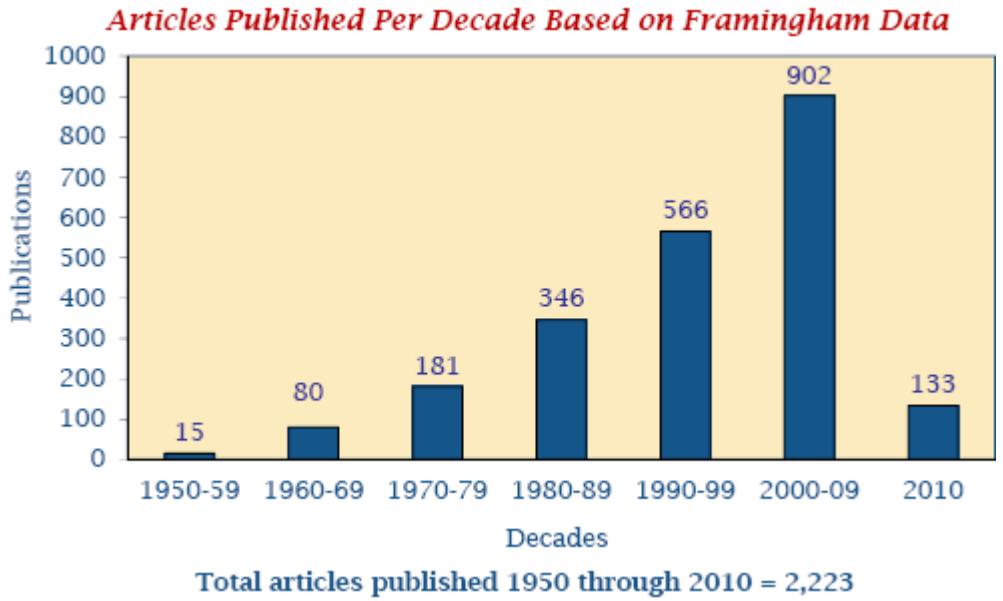
Approximately 5100 residents of this Massachusetts community are followed for > 30 years.
Selected because of a number of factors has permitted assessment of the effects of a wide variety of factors on the risk of numerous diseases

- **stable population,**
- **had a number of occupations and industries represented**
- **had a single, major hospital that was utilized by the vast majority of the population**
- **prepared annually updated population lists that would facilitate follow-up,**

Diseases studied included:

- **coronary heart disease**
- **rheumatic heart disease**
- **congestive heart failure**
- **angina pectoris**
- **intermittent claudication**
- **stroke**
- **gout**
- **gallbladder disease**
- **a number of eye conditions**

The Framingham Heart Study



http://www.framinghamheartstudy.org/risk/index.html

- http://www.ajconline.org/article/S0002-9149(00)00726-8/abstract

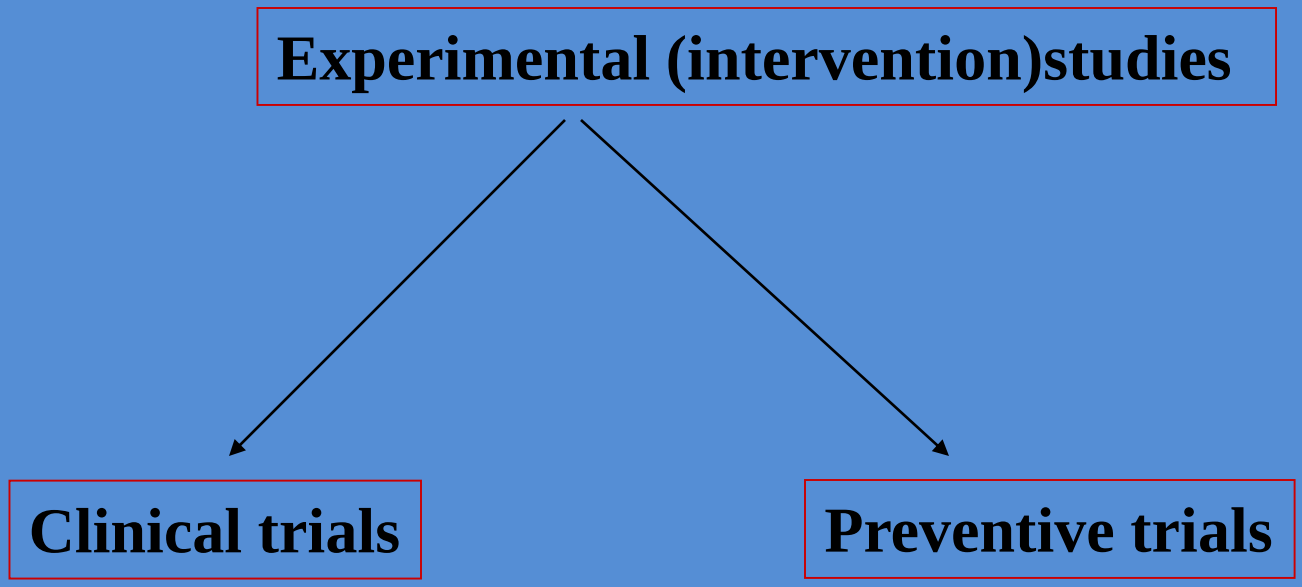
COHORT STUDY DESIGN: Summary

- **In general, can investigate the effect of only a limited number of exposure**
- **Useful for investigating a range of outcomes associated with only one exposure**
- **Useful for study of rare exposure**
- **Not suitable for the study of rare diseases**
- **Follow-up studies are often large and expensive**
- **May take many years to complete**
- **Cannot test current hypotheses**
- **Can measure disease incidence**

Experimental Study Design

A study in which a population is selected for a planned trial of a regimen, whose effects are measured by comparing the outcome of the regimen in the experimental group versus the outcome of another regimen in the control group.

Experimental studies (Intervention)



Experimental Study Design

Different from observational designs by the fact that there is manipulation of the study factor (exposure), and randomization (random allocation) of subjects to treatment (exposure) groups.

Why experimental study design?

- Limitations of theory
- Previous disasters

Clofibrate:

Successfully lowers cholesterol

Treated group: reduced CHD incidence, but higher all causes mortality

- Spontaneous improvements
- Importance of small effects

Clinical trials

- **Individuals with particular disease are randomly allocated into experimental or control groups. randomization is used to ensure that both groups are comparable with respect to all other factors except for the one under investigation.**

- **The experimental group is given the agent being tested and the control group is given either an agent in current use or a placebo**

- **Ideally both patients and the observers should be 'blind' to the treatment being given. This in order to reduce bias.**

Clinical trials

- **Are studies of the effect of a specific treatment on patients who already have a particular disease**

- **They are used to evaluate the efficacy of a preventive or therapeutic agent in the treatment or prevention of a disease**

- **"The most definitive tool for evaluation of the applicability of clinical research" - 1979 NIH release.**

Clinical trials

- **Assessment of each subject must involve bias free accurate measure of outcome**

- **Both groups are followed over a defined period of time when the outcome is then measured in both groups.**

What trials assess

- Drugs
- Surgery
- Type of management
- New services

Why Clinical Trials?

1. **Most definitive method to determine whether a <u>treatment is effective</u>.**

-Provide stronger evidence of the effect (outcome) compared to observational designs, with maximum confidence and assurance

- Other designs have more potential biases
- One cannot determine in an uncontrolled setting whether an intervention has made a difference in the outcome.
- Correlation versus causation

Example: trials of hormone replacement therapy in menopausal women found no protection for heart disease, contradicting findings of prior observational studies

Examples of False Positives

1.High cholesterol diet and rectal cancer
2.Smoking and breast cancer
3.Vasectomy and prostate cancer
4. Red meat and breast cancer
6.Drinking water frequently and bladder cancer
7.Not consuming olive oil and breast cancer

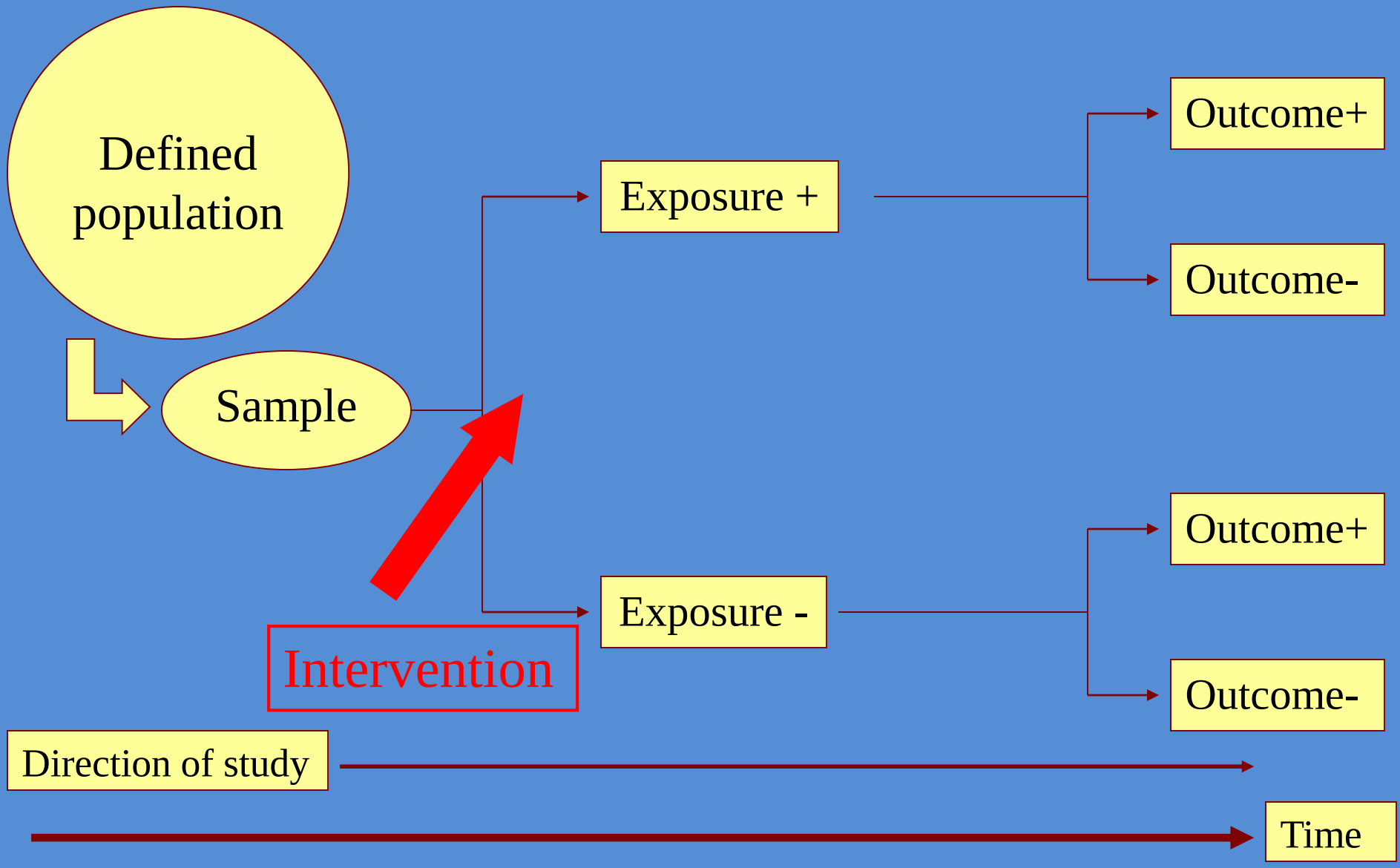
Replication of observational studies may not overcome confounding and bias

Why Performed ?

2. **Determine whether experimental treatments are safe and effective under "controlled environments" (as opposed to "natural settings" in observational designs), especially**

   when the margin of expected benefit is doubtful / narrow (10 - 30%)

Clinical trial



RCT Disadvantages

- Large trials (may affect statistical power)
- Long term follow-up (possible losses)
- Compliance
- Expensive
- Public health perspective ?
- Possible ethical questions
- As above, may take a long time.
- Must be ethically and laboriously conducted.
- Requires treatment on basis (in part) of scientific rather than medical factors. Patients may make some sacrifice

Clinical trials: choice of Design

Depends on:

- Research Questions
- Research Goals
- Researcher Beliefs and Values
- Researcher Skills
- Time and Funds

Clinical trial: Study design

It is also related to:

- **Status of existent knowledge**
- **Occurrence of disease**
- **Duration of latent period**
- **Nature and availability of information**
- **Available resources**

Preclinical

- **Biochemical and pharmacological research.**
- **Animal Studies**

Consists of animal studies that determine the toxicity and bioavailability of a drug. Studies involving animal matrices such as rabbit serum, monkey urine, dog or rat plasma, are all examples of preclinical studies.

Phase I Trials

- **Clinical pharmacology- when the drug is given to healthy people estimate toxicity rates using few (~ 10 - 40) healthy subjects.**

The primary objectives of phase I clinical investigation are:

- **Determine the metabolism and pharmacologic activities of the drug in humans**
- **Side effects associated with increasing doses**
- **Early evidence on effectiveness**
- **Obtain sufficient information about the drug's pharmacokinetics and pharmacological effects to permit the design of well-controlled and scientifically valid phase II clinical studies.**

Phase I Design Strategy

- Designs based largely on tradition
- Typically do some sort of dose escalation to reach maximum tolerated dose (MTD)
- Has been shown to be safe and reasonably effective
- Dose escalation often based on Fibonacci series
  - 1 2 3 5 8 13 . . . .

Phase II Trials

- **Initial clinical assessment: determines whether a therapy has potential using a few very sick patients.**

The primary objectives of phase II studies are:

- **Identify accurately the patient population that can benefit from the drug.**
- **Evaluate the effectiveness of a drug based on clinical endpoints for a particular indication.**
- **Determine the dosing ranges and doses for phase III studies**
- **Common short-term side effects**
- **Risks associated with the drug.**

Phase III Trials

Rigorous testing: large randomized controlled, possibly blinded, experiments

The primary objectives of phase III studies are:

- **Gather an additional information about effectiveness and safety needed to evaluate the overall benefit-risk relationship of the drug.**
- **provide an adequate basis for physician labeling**

Phase IV Trials

- **Post-marketing surveillance: a controlled trial of an approved treatment with long-term follow-up of safety and efficacy.**

The primary objectives of phase IV studies are:

- **Provide additional details required to learn more about a drug's efficacy and/or safety profile.**
- **Study new age groups, races, and other type of patients.**
- **Detect and define of previously unknown or inadequately quantified adverse reactions and related risk factors.**

Types of Clinical Trials

- Randomized
- Non-Randomized
- Single-Center
- Multi-Center
- Phase I, II, III, IV Trials

Purpose of Control Group

- To allow discrimination of patient outcomes caused by test treatment from those caused by other factors
  - **Natural progression of disease**
  - **Observer/patient expectations**
  - **Other treatment**
- Fair comparisons
  - **Necessary to be informative**

Randomized allocation

- Like tossing a coin
- Avoids choosing
- Permits fair comparison

Randomized Controlled Clinical Trial

- Reference: Byar et al. (1976)

  New England Journal of Medicine

- Patients assigned at random to either treatment(s) or control

- Considered to be "Gold Standard"

Ethics of Randomization

- Statistician/clinical trialist must sell benefits of randomization
- Ethics ⇒ MD should do what he thinks is best for his patient
  - **Two MD's might ethically treat same patient quite differently**
- Chalmers & Shaw (1970) *Annals New York Academy of Science*

1. **If MD "knows" best treatment, should not participate in trial**
2. **If in doubt, randomization gives each patient equal chance to receive one of therapies (i.e. best)**
3. **More ethical way of practicing medicine**

- **Bayesian Adaptive designs → More likely assign "better" treatment**

Ethical imperatives

- Must be real doubt
- Obtain inform consent
- Preserve clinical freedom

Defining the patients

- Diagnostic features
- Eligibility criteria (inclusion and exclusion)

Assessing the outcome

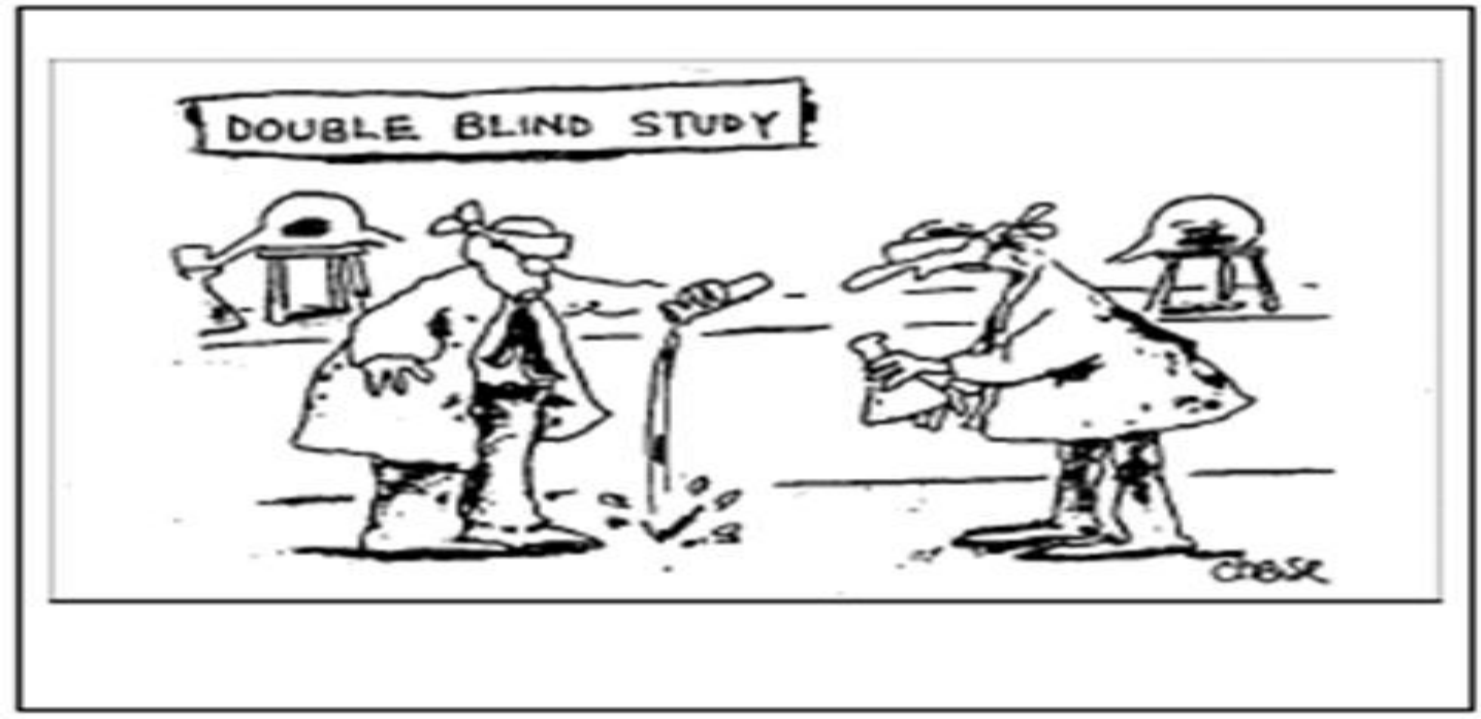
- Clinically relevant
- Easily measured
- Accurately measured

Types of outcomes

- Death
- Clinical measurement
- Symptoms
- Quality of life
- Psychological wellbeing

The need for blinding

- Open
- Single blind
- Double blind
- Triple blind


DOUBLE BLIND STUDY

Definitions

- **Single Blind Study**: A clinical trial where the participant does not know the identity of the treatment received
- **Double Blind Study**: A clinical trial in which neither the patient nor the treating investigators know the identity of the treatment being administered.
- **Triple Blind study: Biostatisticians is also blinded**

Definitions

- **Placebo:**
  - Used as a control treatment

    1. An inert substance made up to physically resemble a treatment being investigated

    2. Best standard of care if “placebo” unethical

    3. “Sham control”: Faked surgical intervention with the patient's perception of having had a regular operation

Definitions

- **Adverse event:**
  - An incident in which harm resulted to a person receiving health care.
  - **Examples:** Death, irreversible damage to liver, nausea
  - Not always easy to specify in advance because many variables will be measured
  - May be *known* adverse effects from earlier trials

Surrogate Endpoints

- Response variables used to address questions often called endpoints
- Surrogates used as alternative to desired or ideal clinical response to save time and/or resources
- Examples
  - Suppression of arrhythmia (sudden death)
  - T4 cell counts (AIDS or ARC)
  - Cholesterol (heart disease)
- Often used in therapeutic exploratory trials
- Use with caution in confirmatory trials

Summary of trial design

- Specify the treatment
- Define study group
- Random allocation
- Blinded outcome assessment
- Fair interpretation

Clinical trial

Common problems

- Too few patients
- Failed randomization
- Patients lost to follow-up
- Flawed analysis-interpretation
- Power of study: not big enough

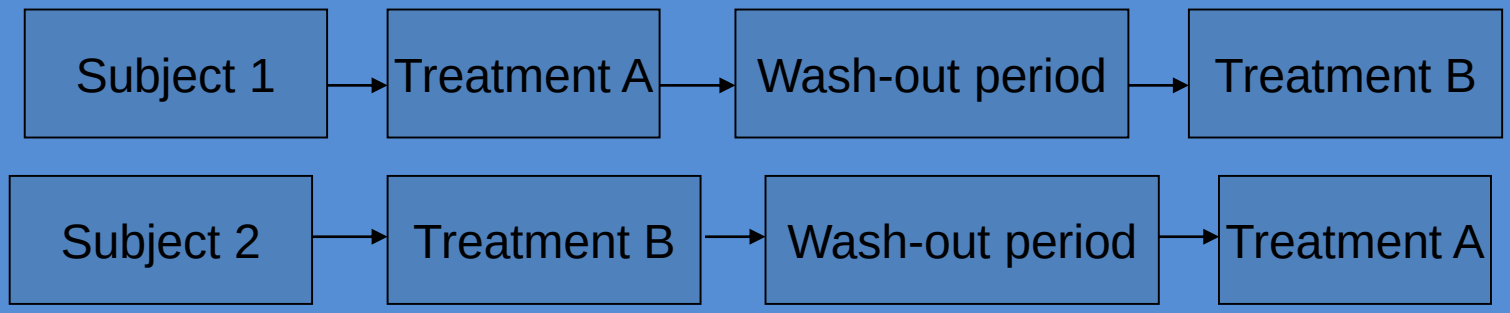
Cross-over clinical trial

Each patient gets both treatments

Half get A then B

Half get B then A

Wash-out period in between



Cross-over clinical trial

- Cross-over design
- Patient as own control

-Reduce variations

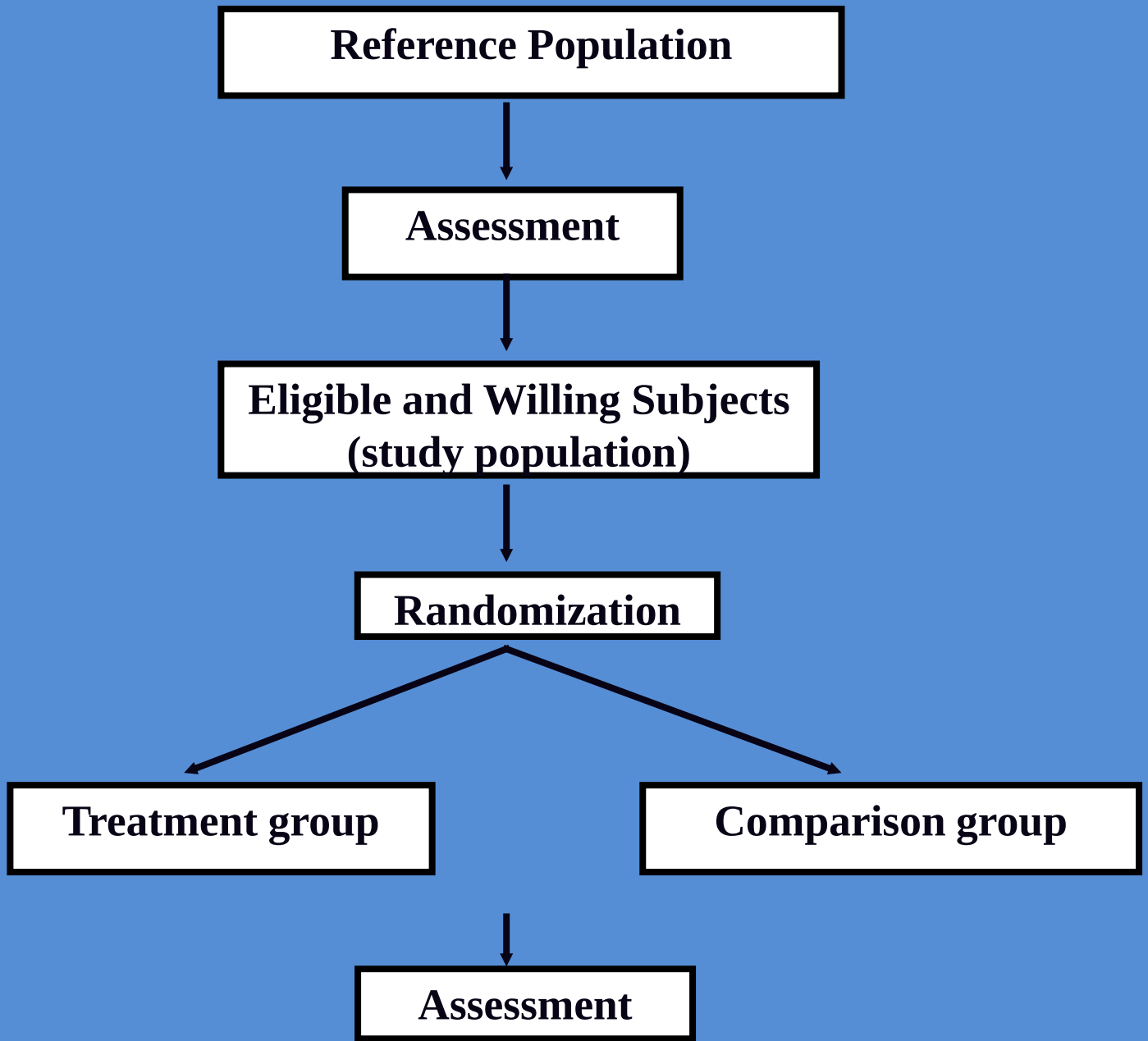
-Much smaller sample size

Requirements: Carry over period(s)

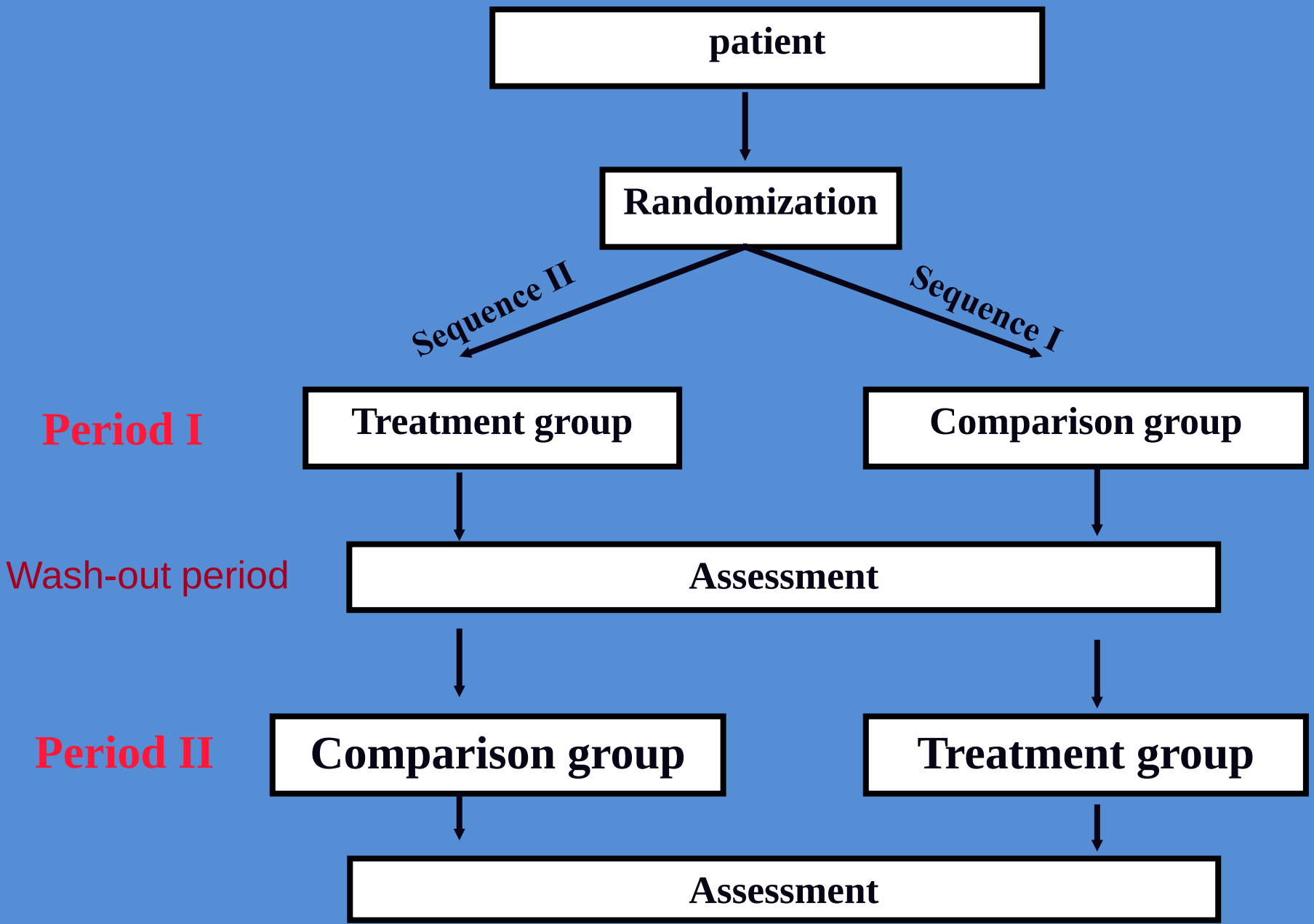
Key elements of RCTs

- Selection of subjects
- Comparison group
- Randomization
- Allocation of treatment
- Blinding (single, Double blind design/placebo)
- Intention to treat analysis in which the treatment and control groups are analyzed with respect to their random allocation, regardless of what happened subsequently
- Ethical considerations

Parallel Design



Crossover Design



Preventive trials

Are studies of the effect of a possible preventive measure on people who do not yet have a particular disease. Another type of preventive trial is a study of the effect of a possible preventive measure on whole communalities.

Preventive trials

- **The risk of developing any particular disease among the people who are free from disease is small. Because of this, preventive trials usually require a greater number of subjects than clinical trials, and are therefore more expensive**

- **This expense limits their use to the study of preventatives of extremely common or extremely severe diseases**

  e.g. vaccination to prevent whooping cough

  vaccination to prevent poliomyelitis

- **When a disease occurs rarely, it is more efficient to study those people thought to be at high risk of disease , e.g. vaccine to prevent Hepatitis B**

Preventive trials

- As in clinical trials, the preventatives should be given so that the individuals who do and do not receive the preventative are as comparable as possible. This is often difficult.

- In some types of trials the preventative have to be administered to communities rather than individuals, e.g. water fluoridation to prevent dental caries

Results of a trial to determine whether A vaccine could prevent whopping cough

	No. with Whooping cough	No. without Whooping cough
Number vaccinated 3801	149(4%)	3652(96%)
Number not vaccinated 3757	687(18%)	3070(82%)

Community Trials

- A community participates in a behavioral intervention, nutritional intervention, a screening intervention, etc
- Intervention: Any program or other planned effort designed to produce changes in a target population.
- *Community* refers to a defined unit, e.g., a county, state, or school district.
- Communities are randomized and followed over time.
- Determine the potential benefit of new policies and programs.

Examples:

- A community-level intervention for tobacco control might combine a school curriculum for youth to prevent initiation of smoking
- A media campaign aimed at reducing smoking rate

Sampling techniques

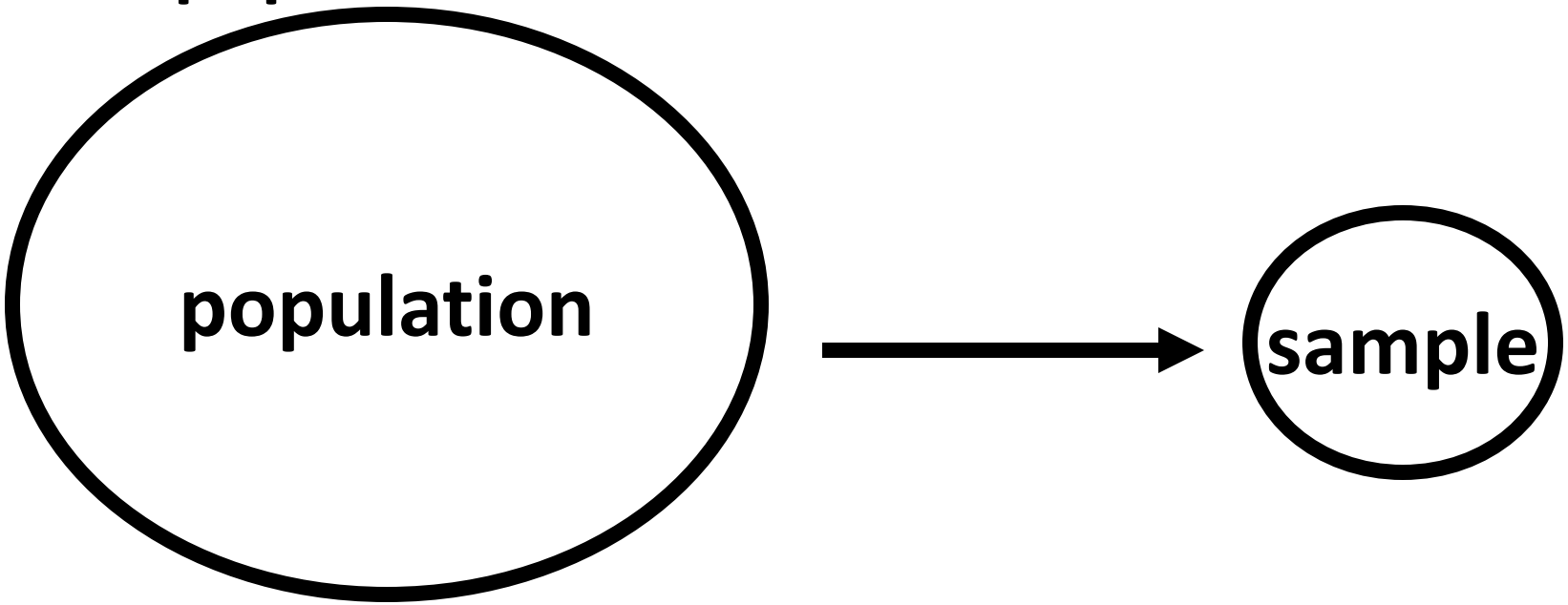
Dr Munir Abu-Helalah
MD MPH PHD
Associate Professor of Epidemiology and Preventive Medicine

Sampling techniques

- A Word About Sampling…
- The **population is all the members of** the group you are researching (e.g., all youth in our city)
- The **sample is the selection of the** population who will be asked questions
- To **generalize is to state that what you**say about your sample can be applied to the rest of the population

Sampling

Sampling is a process by which we study a small part of a population to make judgments about that population.



Selection of samples

Types of sampling most frequently used in health surveys

- Complete or comprehensive survey of each unit in the population (e.g. nurses in a single hospital)

- Probability sample survey
  - Systematic sampling
    - Record reviews
    - Studies of health care workers
  - Cluster sampling
    - Used in surveys of widely dispersed populations

Definitions

A study unit may be a person, a health facility, a prescription, or other such unit.

The study population, sometimes called the reference population, is the collection of the entire population of all possible study units. Again, this population may be people, health facilities, prescriptions or other such units.

A representative sample has all the important characteristics of the population from which it is drawn.

SAMPLING METHODS

A sampling frame is a list of all of the available units in the study population. If a complete listing is available, the sampling frame is identical to the study population. The method of sampling depends on whether there is a sampling frame available. If a sampling frame exists, or if it can be created, probability sampling is used. If there is none available, probability samplings cannot be used.

non-probability sampling

using non-probability methods is likely to be less representative than a probability sampling and so study results are less valid.

NON -PROBABILITY SAMPLING METHODS

1. Convenience Sampling

is a method by which, for convenience sake, the study units that happen to be available at the time of data collection are selected in the sample. This is the least representative sampling method.

NON -PROBABILITY SAMPLING METHODS

Quota sampling

is a method by which different categories of sample units are included to ensure that the sample contains units from all these categories. For example, a quota sample of patients from a health center that might included 10 patients with diabetes, 10 with diarrhea, and 10 with malaria.

Quota sampling is a method of sampling widely. Interviewers are each given a quota of subjects of specified type to attempt to recruit for example, an interviewer might be told to go out and select 20 adult men and 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing.

It suffers from a number of methodological flaws, the most basic of which is that the sample is not a random sample and therefore the sampling distributions of any statistics are unknown.

Types of Probability Samples

- **Simple Random**
- **Systematic Random**
- **Stratified Random**
- **Random Cluster**
- **Stratified Cluster**
- **Complex Multi-stage Random**

(various kinds)

1. Simple Random Sampling-1

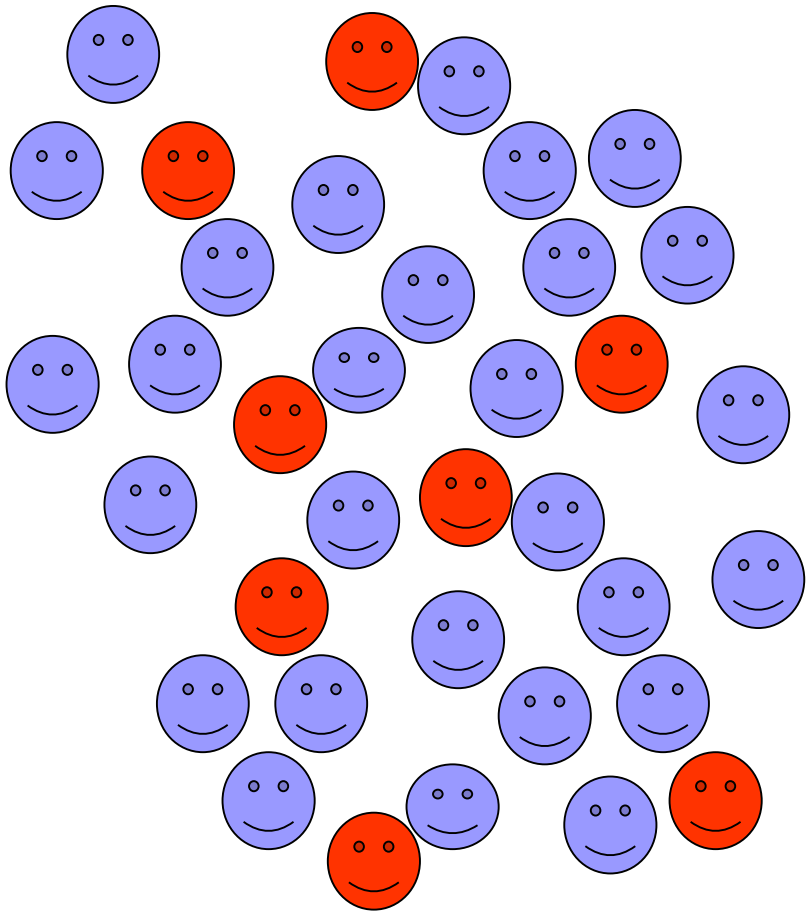
a) Make a numbered list of all units in the reference population from which you will select the sample (for example, a list of all the health centers in the country).

b) Decide on the size of the sample (say 20 facilities).

c) Choose the facilities to include by a lottery method. (For example the numbers of all the facilities can be placed in a box and drawn, a random number table can be used, or random numbers can be generated using a spreadsheet or calculator.)

Simple Random Sampling-2

- **Each element in the population has an equal probability of selection AND each combination of elements has an equal probability of selection**
- **Names drawn out of a hat**
- **Random numbers to select elements from an ordered list**



How to select a random number?

- Flip a coin
- Choose a number from a ‘hat’
- Bank note
- Calculator
- Computer
- Table of random number

2. Systematic Sampling

In systematic sampling, sample units are selected from a numbered list of all units in the study population by using a regular interval, starting from a random sampling starting point.

To calculate the sampling interval,

- **Determine the total number of units in the population**
- **Determine the sampling interval**

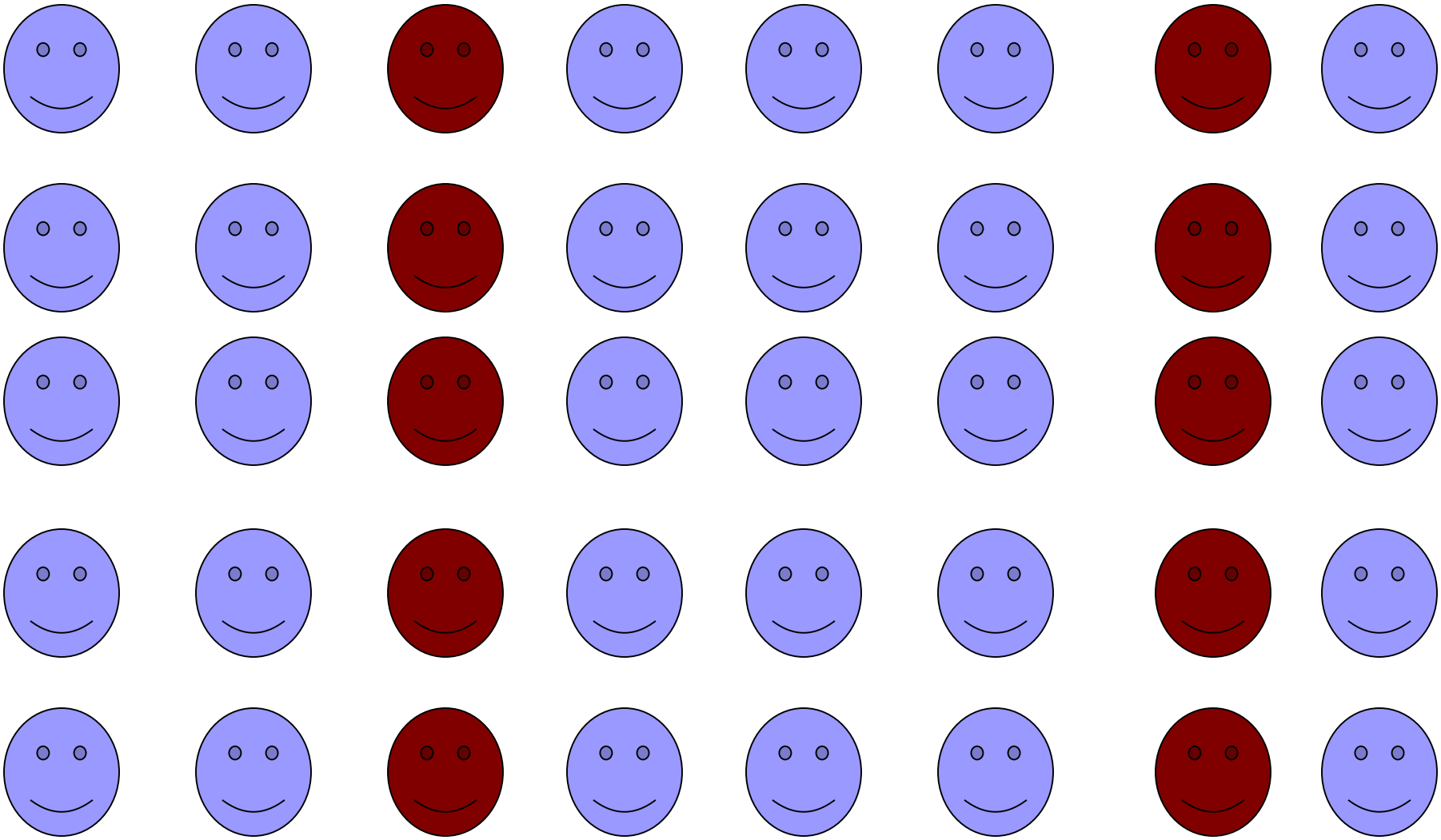
$$\text{sampling interval} = \frac{\text{number of units}}{\text{desired sample size}}$$

For example, if we want to select 20 health centers from a list of 46 in our sampling frame, our sampling interval would be 46/20 = 3.

The first facility chosen in this case can be 1, 2 or 3, which are all the possible sampling units within the first sampling interval. This is selected by choosing a random number with one digit less than or equal to the sampling interval.

Later facilities are selected by adding the sampling interval to the previous result. If the first result was 3. then the next facilities selected would be facility 6, 9 and so forth. The method just described gives every unit an equal chance of being selected.

Sample 12, sampling intreval=48/12=4



Example:

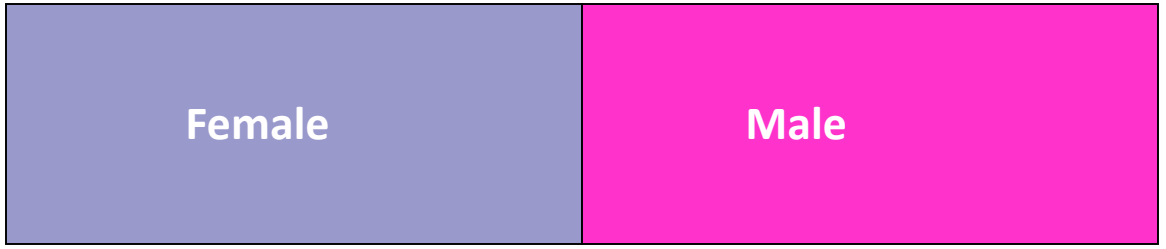
Assume you are doing a study involving children under 5. There are a total of 1500 households, and you have a required sample size of 100 children. From a preliminary study you have done, there is one child every 2.5 households. Therefore you would need to visit 100X2.5 or 250 households to find the required 100 children.

sampling interval=1500/250=6 (Visit every 6^{th} household)

- **select a number between 1 and the sampling interval**
- **add the sampling interval to the chosen starting point to obtain the second sampling unit, add the interval to the second unit ….**

3. Stratified Sampling-1

- Stratified sampling is used when the reference population contains clearly different sub-populations, which should be considered separately.



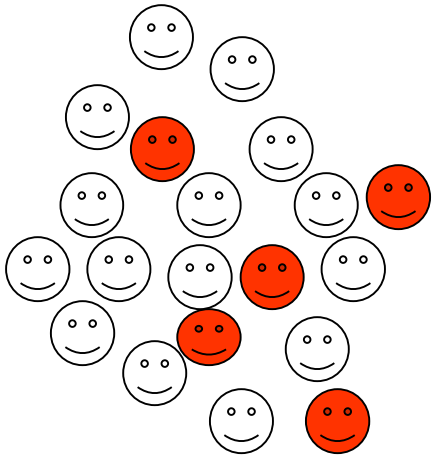
- When stratified sampling is used, the sample frame (the list of the overall population) is sorted into two or more groups. These different strata (groups) may then be sampled either randomly or systematically.

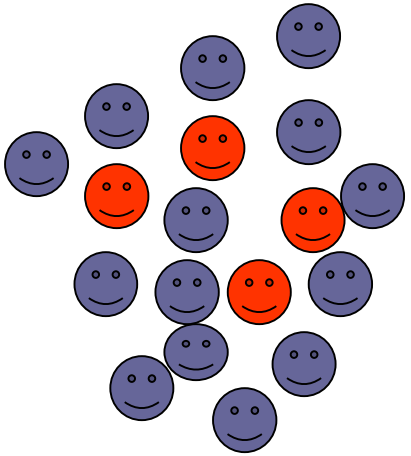
- Basis for grouping must be known before sampling

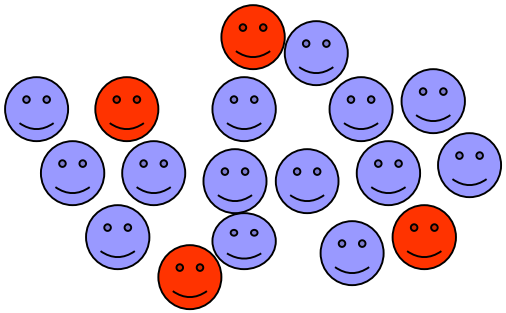
- Select random sample from within each group

Stratified Sampling-2

- For a given sample size, reduces error compared to simple random sampling IF the groups are different from each other
- Tradeoff between the cost of doing the stratification and smaller sample size needed for same error
- Probabilities of selection may be different for different groups, as long as they are known
- Over sampling small groups improves intergroup comparisons







4. Cluster Sampling

- **Cluster sampling: Dividing the population into subgroups called clusters (not as homogeneous as strata), randomly sampling clusters, and then possibly selecting a random sample of people in each cluster.**

- **In a cluster sample, a group of sample units is selected together, rather than each unit being selected separately.**

(Sampling unit is a group of individuals) e.g.

- **Households**
- **Health centers**
- **Schools**
- **Village**

- **Selection with probability proportional to size**

e.g. EPI WHO sampling procedure of selecting 30 groups of 7 children is a common cluster sampling method.

- **The main advantage of cluster sampling is that the method is easy to use and often logistically simpler to organize.**
- **The disadvantage is that the samples selected may be less representative especially when the number of clusters selected is small. As a rough guide, double the sample size if cluster sampling is used.**

Steps in selecting a cluster sample

- Enumerate all population concentrations in the sampling universe
- Draw up a cumulative population list
- Determine the sampling interval
- Pick a number between 1 and the sampling interval from a random number table
- Add the sampling interval to the chosen starting point to obtain the second cluster, add the interval to the cluster
- Select individuals=sample size/number of clusters
- Design effect

Stratification vs. Clustering

Stratification

- Divide population into groups different from each other: sexes, races, ages
- Sample randomly from each group
- Less error compared to simple random
- More expensive to obtain stratification information before sampling

Clustering

- Divide population into comparable groups: schools, cities
- Randomly sample some of the groups
- More error compared to simple random
- Reduces costs to sample only some areas or organizations

5. Multistage Sampling

In multistage sampling, the methods described above can be combined. For example, we might wish to select 32 health facilities in a country containing 56 districts, each of which contains a number of health facilities. From the 56 districts, 16 districts would first be selected. In each district two health facilities would then be randomly selected. This would be two-stage random sampling.

Surveys

A survey may be defined as a collection of information from all individuals or a sample of individuals chosen to be representative of the population from which the are drawn

Types of information collected by sur

- Morbidity prevalence
- Mortality
- Detailed risk factors or behavioral information
- Knowledge, attitudes, and practices
- Physical signs (paralysis, splenomegaly, malnutrition)
- Serological or laboratory tests

Characteristics of survey

- representative if sample chosen correctly
- Single point in time –snapshot
- Provide more in depth information than surveillance or chart reviews
- Usually performed by a limited number of personnel specially trained to perform surveys
- Can sometimes be expensive, time consuming to perform
- Cannot be used to monitor change unless repeated at a later time therefore may be better for situational analysis than for ongoing monitoring of a problem or a programme

When to do a survey

- When accurate population-based data are needed to determine the magnitude of the problem
- When more detailed or recent information is needed than is available from record review or surveillance (demography, examination, laboratory)
- When information is needed on health problems that may not routinely be seen by health providers
- When information is needed on health behaviors or health knowledge and attitudes not routinely available through existing mechanisms

Survey

Key Concepts of survey design:

1. Primary data
2. Communication
3. Sample
4. Representative

TYPE OF MEASUREMENT

- *Attitudes*: What people feel
- *Knowledge*: What people know
- *Beliefs*: What people think is true: their beliefs
- *Behaviours*: What people do or have done
- *Evaluation*: Peoples perception of thing are/were

Classifying Survey Research Methods

1. By method of communication.
   a) Personal Interviews
   b) Telephone interviews
   c) Self-administered interviews
2. By time frame (Temporal classification).
   a) Cross-sectional surveys
   b) Longitudinal surveys

Temporal Classification of Survey Research

1. ***Cross-sectional studies:*** studies in which various segments of a population are sampled and data collected at a single point in time.
2. ***Longitudinal studies:*** studies in which data are collected at different points in time

Survey

- **Focus on personal and demographic characteristics, illness and health related habits**
- **These surveys may also examine frequencies of disease and other characteristics may be examined in relation to age, sex, location, education, etc**

Range of uses of survey

Target groups:

1. Patients

Examples of topics of interest:

Need for services

Satisfaction with care given

Side effects of care

Compliance with therapy

Quality of life

Health behaviour and beliefs

Range of uses of survey

Target groups:

2. Health professionals

Examples of topics of interest:

Knowledge and experience

Activities undertaken

Attitudes to the provision of care

Sources of stress and dissatisfaction

Educational needs

Range of uses of survey

Target groups:

3. Relatives and carers

Examples of topics of interest:

Understanding of illness and its treatment

Satisfaction with information given

Knowledge of available support services

Attitudes to and stresses of caring

Range of uses of survey

Target groups:

4. General public and selected subgroups

Examples of topics of interest:

Morbidity

Quality of life

Unmet need for services

Access to services

Use of preventive services

Health behaviour and beliefs

Range of uses of survey

Target groups:

5. Health care facilities

Examples of topics of interest:

Availability of equipment

Staffing levels

Training and experience of staff

Extent of provision of services

Nature of service organisation

Align your measure with your theoretical orientation

- Good survey measures must be grounded on sound theory and conceptual definitions

Examples:

- Health Belief Model
- Illness behaviour model

ADVANTAGES OF SURVEYS

1. **Can complete structured questions with many stakeholders within a relatively short time frame.**
2. **Can be completed by telephone, mail, fax, or in-person.**
3. **It is quantifiable and generalizable to an entire population if the**
4. **population is sampled appropriately.**
5. **Standardized, structured questionnaire minimizes interviewer bias.**
6. **Tremendous volume of information can be collected in short period of time.**
7. **Speed: faster data collection than other methods**
8. **Cost: relatively inexpensive data collection**
9. **Accuracy**
10. **Efficiency: measured as a ration of accuracy to cost, surveys are generally very efficient data collection methods**

DISADVANTAGES OF SURVEYS

1. More difficult to collect a comprehensive understanding of respondents' perspective (in-depth information) compared to in-depth interviews or focus groups.
2. Survey error: Potentially large sources of error in surveys
3. Communication Problems - Each of the different communication survey methods has its own unique problems.

WHY DO YOU WANT TO DO THIS SURVEY?

1. Why have you chosen to conduct a survey? What did you want to learn from the results and/or what decisions need to be made from the results?

Clearly write down your survey research questions.

2. When considering why you want to do this survey?

Be very specific

3. Focus on the ‘need’ to knows, not the ‘nice’ to know

Does your reasoning fit the uses of surveys?

If not, perhaps you should consider a different method.

Survey design

4. WHO ARE THE STAKEHOLDERS?

- Stakeholders are all those individuals who would have an interest in the questions you are asking and the results obtained (i.e. Stakeholders of the screening program/service/medication)

Survey design

5. WHO IS THE POPULATION OF INTEREST?

- Describe the population you are interested in surveying:
- What is their demographics (age, gender, ethnicity)?
- Where do they live?
- Are they all very similar or are there unique differences?
- Are you interested in any sub-groups of this population?
- Determining the characteristics of your population of interest gives you some indication of:

I. How you can get a sample of respondents
II. Whether you need to set quotas for subgroups

Survey design

6. What is the best way to reach them?

What is the best way to communicate with them?

Medium (phone, fax, mail, e-mail)

Time of day

Time of week

Survey design: summary

- Describe the group of interest
- Obtain a list of possible participants
- Decide on sample size
- Select the method sampling

Developing research question

- Having the idea
- Identify the research question
- Review the literature
- Dissect out the research question(s)
- Select the key question
- Refine the research question

Major steps in surveys

- determine what information is needed to establish objectives and draw up the table shells you will be using to analyze the data
- Determine
  - Sampling universe (what is the population you are sampling from?) geographic area to be sampled, as well as who in the population is to be included in the survey (children under 5, women of reproductive age, clinic Attendees at a certain clinic) need to be determined
  - Sample size
  - Sampling method

Major steps in surveys

- Establish the methodology for collecting the information (questionnaires, day-to-day methods,plans for specimens, etc).
- Train the staff- field test questionnaire & other data collection methods
- Draw survey maps, arrange logistics and supervision
- Conduct survey
- Organize, analyze and write report in a timely fashion

Steps in conducting a study

Step 1: Determine the objectives of your study

- **Why conduct a Survey?**

Who is the population of interest?

- **What issues need to be explored?**
- **What question(s) are you trying to answer?**
- **Who will be using your findings?**
- **How will these findings be used?**

Step 2: Determine the exposure and outcome variables and decide how you will define them

- **Sources:literature, experts, focus groups, preliminary interviews**
- **Be able to justify the inclusion of each variable**
- **Avoid temptation to include variables that "might be interesting"**
- **Realize you may need more than one study**
- **Decide how variables to be classified**

Step 3: develop preliminary “skeleton” tables

- Begin with simple descriptive characteristics
- Develop shells for two way tables
- Develop shells for any stratified tables

Step 4: determine

- Who will be the study subjects
- Methodology
- Sample size

Step 5: design a questionnaire

- Decide on what questions to ask
- Set the types of response formats
- Set the layout of the questionnaire
- To fill in the blanks in the skeleton tables
- The analysis should drive the questionnaire rather than vice-versa!
- Pilot your questionnaire

Step 6: Establish a sampling plan for data collection and work out the logistics

- establish the methodology for collecting the information,
  - Types of questionnaire
  - Day-to-day methods
  - Plans for specimens

Step 7: Determine the personnel needs

- Types of people and necessary person-hours
- Develop appropriate descriptions of responsibilities for each level of personnel
  - Supervisors
  - Surveyors
  - Drivers/guides
  - translators

Step 8: Field test the questionnaire in the population in which it is to be used and determine whether there are operational problems

- Revise the questionnaire / methods
- Develop other necessary forms
  - record-keeping forms for interviewers to keep track of sites visited

Step 9: Develop instruction manuals for survey personnels

- To detailing how questionnaires are to be filled
- How the sample is to be selected
- How field supervision will be performed

Step 10: select and train the personnel to be used to collect the data

- Keys to the training are:
  - Information
  - Examples
  - practice

Step 11:Develop check list of materials needed for field work

- Forms
- Papers
- Pencils
- Clipboards
- Paperclips
- Sleeping bags
- tents
- Per diems
- Payment schedule
- Review of data forms

Step 12: collect the data

- assuring:
  - quality
  - completeness

  (through supervisory visits and review of data forms

Step 13: Edit your data to determine errors in collection, coding, transcription, or data entry

- If field entry, build in edit checks
- Look for abnormal values, unexpected population distribution
- Perform plausibility edits

- Go back to the source whenever possible
- Avoid second-guessing
- Be consistent
- Fix errors as soon as it occurs
- Document the fix

Step 14: do the data analysis

- Calculate the response rates
- Fill out the skeleton tables
- Collapse categories
- Think about what your data means
- Measures of association and statistical tests keeping in mind:
  - Study design (matching, design effect)

Step 15: interpret your data

- Meaning of the results
- Significance testing

Step 16:Writing up

- Immediately
- Disseminate to the appropriate people

A pilot test

- **Is an evaluation of the specific questions, format, question**
- **sequence and instructions prior to use in the main survey.**
- **Pilot testing is a crucial step in conducting a survey. Even modest pretesting**
- **can avoid costly errors.**

Questions answered by the pilot test include:

1. Is each of the questions measuring what it is intended to measure?
2. Are questions interpreted in a similar way by all respondents?
3. Do close-ended questions have a response which applies to all respondents?
4. Are the questions clear and understandable?
5. Is the questionnaire too long?
6. How long does the questionnaire take to complete?
7. Are the questions obtaining responses for all the different response categories or does everyone respond the same?