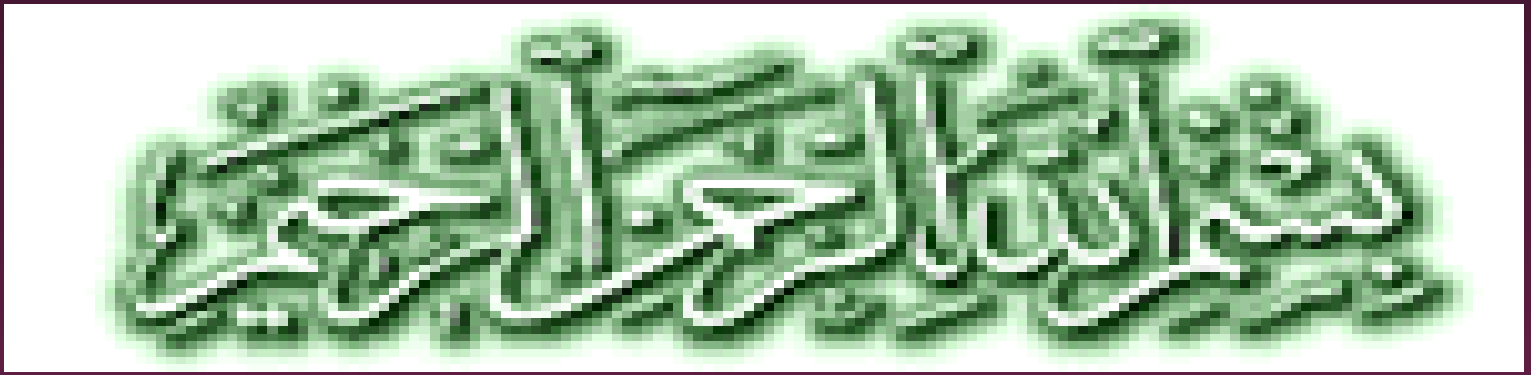
Epidemiological Study Designs

Prof. Lubna Jamal Abdelmalek

Professor Of Family and Community Medicine
Faculty of Medicine



اللهم وفقنا لما تحب و ترضي ولا تكلنا الي انفسنا طرفه عينا ابدا

OUTLINE

- INTRODUCTION
- DEFINITIONS
- CLASSIFICATION
- STUDY DESIGNS
- HOW TO CHOOSE YOUR RESEARCH DESIGN

DEFINITION OF EPIDEMIOLOGY

Epidemiology is derived from the greek terms epi=upon,among. demos=people. Logos study.

Epidemiology is the study of frequency, distribution and determinants of health related state or event in a specified human population and the application of this study to the control of health problem.

EPIDEMIOLOGY [DEFINITION OF KEY TERMS]

Frequency :(including rates & risks) .
Distribution : pattern of health events (person, place, time)
Determinants : factors or events that are capable of bringing about a change in health
Human population : Epidemiology examines health events among population groups rather than individuals.

EPIDEMIOLOGY [DEFINITION OF KEY TERMS]

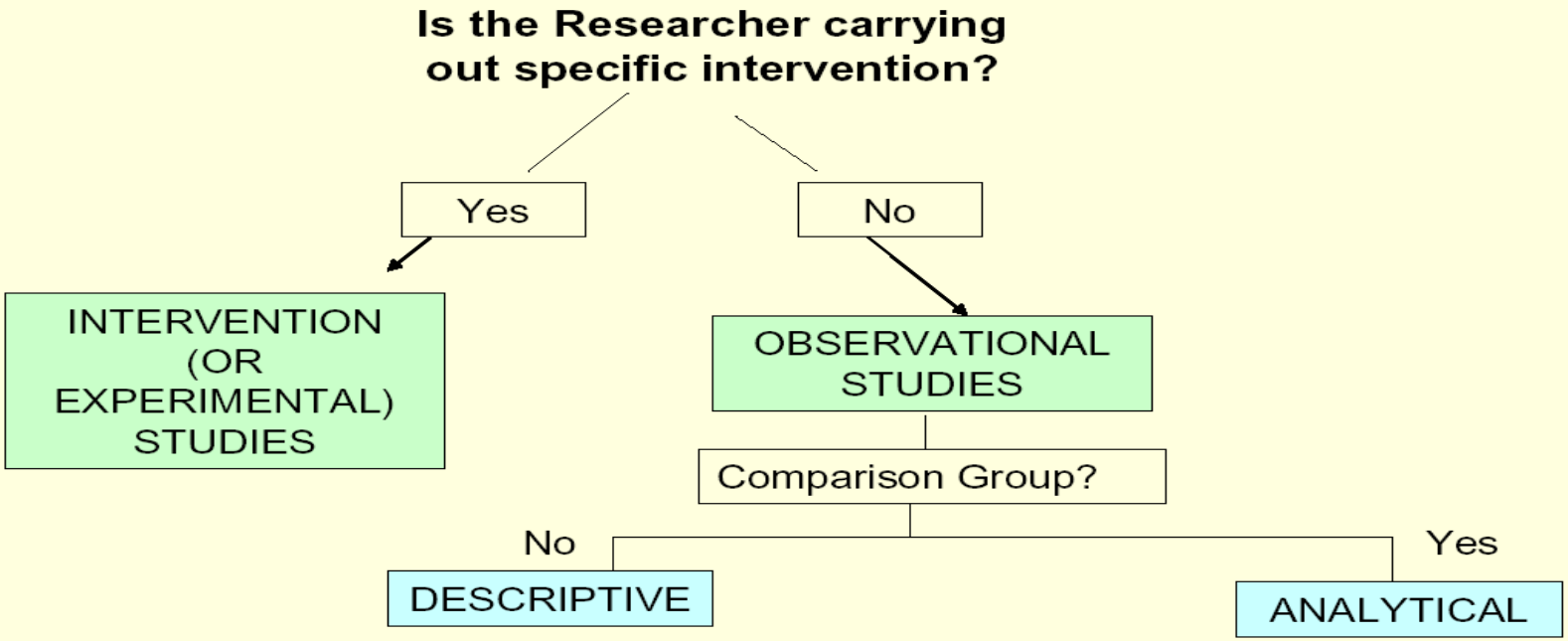
Health related states: infections, chronic diseases & physiological events &various states of health such as disability, injury, mortality

Health related events : immunization, hospital attendance, bed occupancy

Application : basis for directing interventions

CLASSIFICATION

Epidemiologic Study Designs



OBSERVATIONAL VS EXPERIMENTAL STUDIES

- **Observational studies :**

  Allow nature to take its cause; the investigator measures but does not intervene
- Descriptive study: focuses on the description of the occurrence of a disease in a population
- Analytical study analyses relationships between health status and other variables

OBSERVATIONAL VS EXPERIMENTAL STUDIES

- **Experimental or interventional studies:** involve an active attempt to change a disease determinant(e.g an exposure or a behaviour) or the progress of a disaese (through treatment)
- The studies are based on a grp which has had the experience compared with control grp which has not had the experience.

OBSERVATIONAL STUDY DESIGNS

- Descriptive
  - Case report
  - Case series
  - Cross-sectional studies
  - Ecological or correlation studies
- Analytic
  - Case-control studies
  - Cohort studies

Descriptive Epidemiological Studies

Basic Triad of Descriptive Epidemiology

THE THREE ESSENTIAL CHARACTERISTICS OF DISEASE WE LOOK FOR IN DESCRIPTIVE EPIDEMIOLOGY ARE:

- **PERSON**
- **PLACE**
- **TIME**

Descriptive Epidemiological Studies

- Concerned with describing the general characteristics of the distribution of a disease (person, place, time)
  - "Who" or Person
    - Age, Sex, Race, Socioeconomic Status
  - "When" or Time
    - Time of day, week, month, season, year, decades
    - incubation period
  - "Where" or Place
    - country, state, street, urban or rural.

Descriptive Epidemiological Studies

- objective is to describe the patterns and trends
- help to generate hypothesis
- help to plan programs
  - measure frequency of disease or other health outcome (occurrence)
  - measures determinants (risk factors) and effects on health outcomes
  - risk factors and effects may be measured over time

Descriptive Epidemiological Studies

- Used to compare disease frequencies between different groups during the same period of time

  or

  in same population at different time points

- Cannot be used to test hypotheses - cannot link exposure to occurrence of disease in the same person.

Descriptive Epidemiological Studies

- Case report
- Case series
- Cross-sectional studies
- Ecological or correlation studies

Descriptive Studies of Individuals

Case Report:

- careful, detailed reports by 1 or more clinicians of the profile of 1 patients.
- documents unusual medical occurrence and can represent the first clues to the formulation of hypothesis, generally report a new or unique findings and previous un described disease.

Case Report

- What?

  the profile of a single patient is reported in detail by one or more clinicians

- Example

  In 1961, a published case report of a 40 year-old women who developed pulmonary embolism after beginning use oral contraceptive

Case Series

Case series - # of patients

collection of individual case reports which may occur within a fairly short time, and experience of a group of patients with similar diagnosis.

Advantages: Useful for hypothesis generation

Informative for very rare disease with few established risk factors

Disadvantages

Cannot study cause and effect relationships

Cannot assess disease frequency

Case Series

- What?

  Experience of a group of patients with a similar diagnosis.
  - Informative for very rare disease with few established risk factors.

- Example

  Angiosarcoma of liver in three vinyl chloride plant workers.

Case Report and Case Series

- Profile of a clinical case or case series which should:
  - illustrate a new finding,
  - emphasize a clinical principle, or
  - generate new hypotheses
- Not a measure of disease occurrence!
- Usually cannot identify risk factors or the cause (no control or comparison group)

Ecological or Correlation

The whole population is the unit of analysis.

- Measures that represent characteristics of entire populations are used to describe the disease in relation to some factor of interest (such as age, calendar time, food consumption, drug use and utilization of health services)
- aim to show strength of the ecological association.

CORRELATIONAL STUDY DESIGN

ADVANTAGE:

- Compares events among nations

DISADVANTAGE:

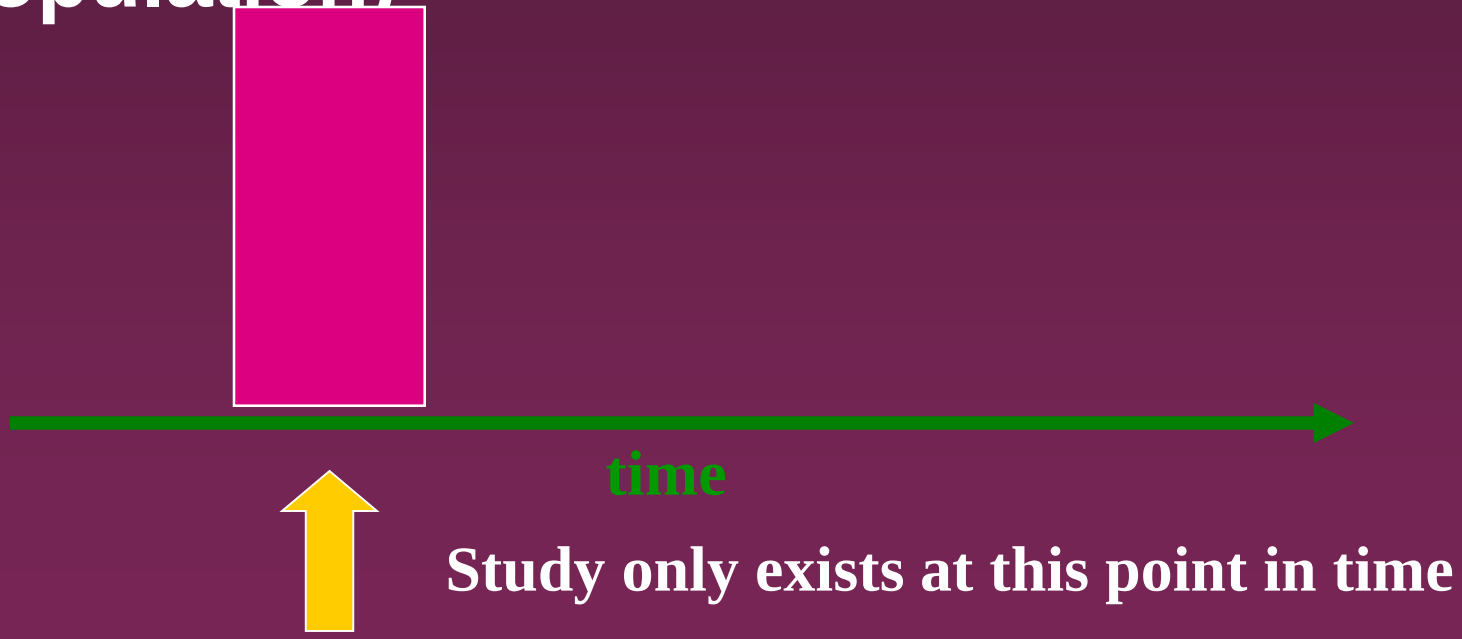
- Doesn’t compare individuals, so it might lead to overgeneralization.

Cross-Sectional Study

- or "PREVALENCE STUDY"
- Risk factors (exposures) and disease outcome are ascertained at a single point in time in a cross-sectional sample of subjects.
- It answers the question "WHAT IS HAPPENING RIGHT NOW?"

Cross-sectional studies

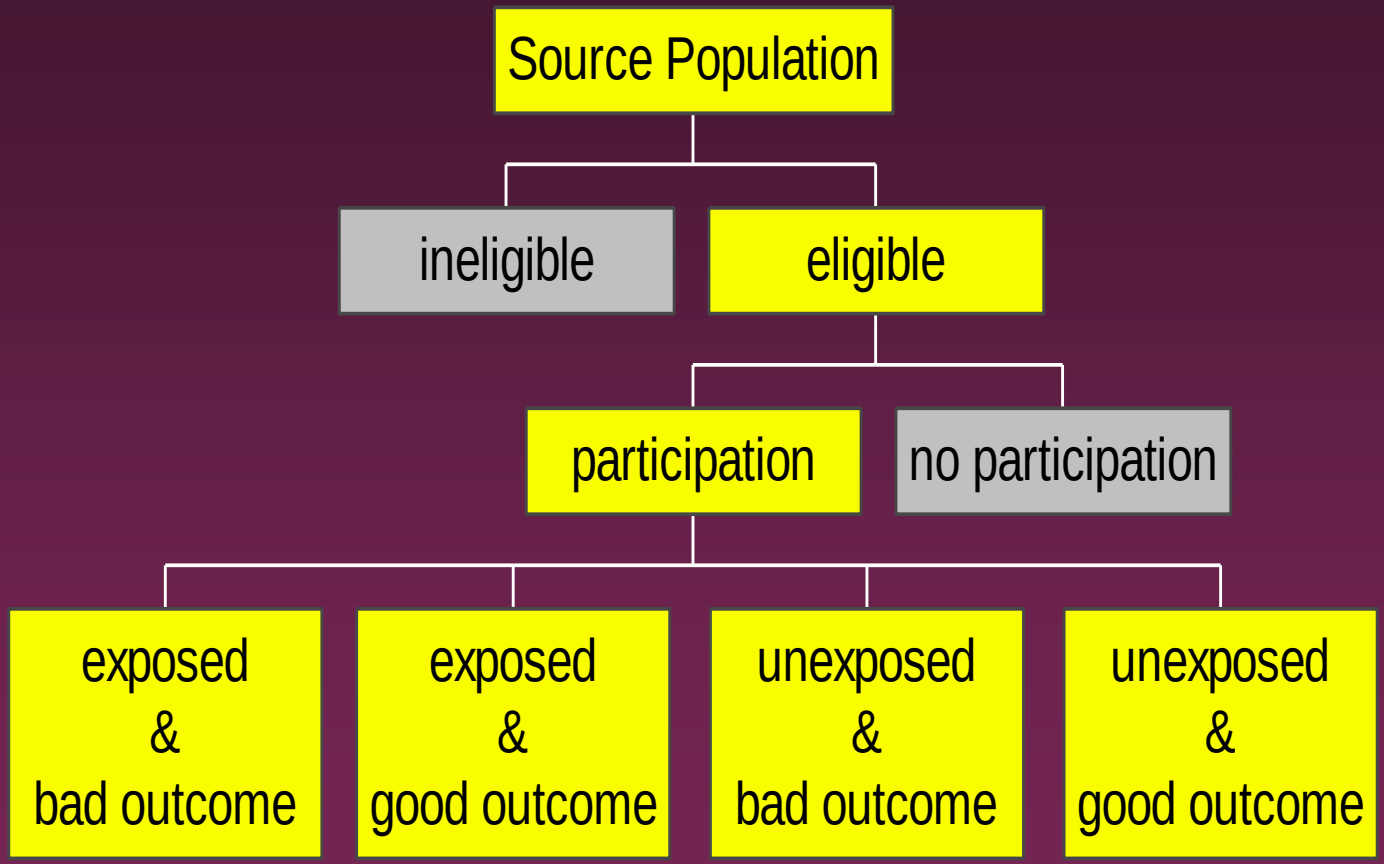
- An “observational” design that surveys exposures and disease status at a single point in time (a cross-section of the population)



Cross-sectional

- often interest is to describe frequency and pattern of either disease or health-related outcome occurrence.So not to test (associations) but to generate hypothesis and describe occurrence.
- existing traits, be it disease or health-related outcome, are measured at same time. Assessment of temporality in found association is not possible
- neither cases nor comparison group, if exist, are pre-selected (post hoc selection).

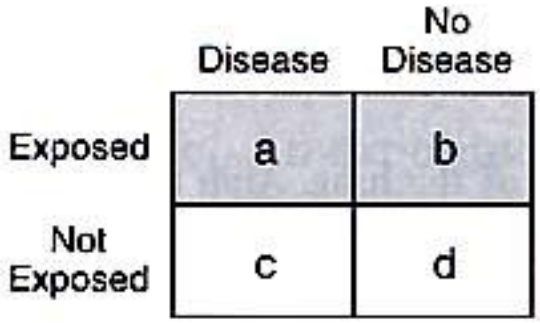
Cross-Sectional Study

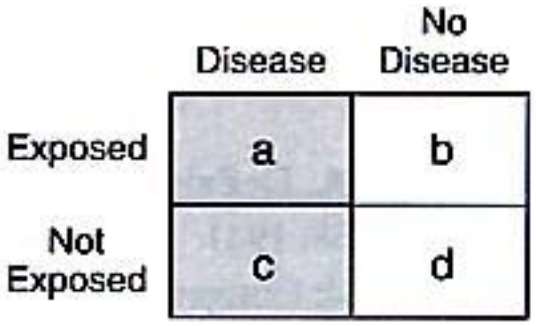


Cross-Sectional Study

	Disease		
Exposure	yes	no	total
yes	a	b	a + b
no	c	d	c + d

Cross-Sectional Study





Prevalence of disease compared in exposed and nonexposed

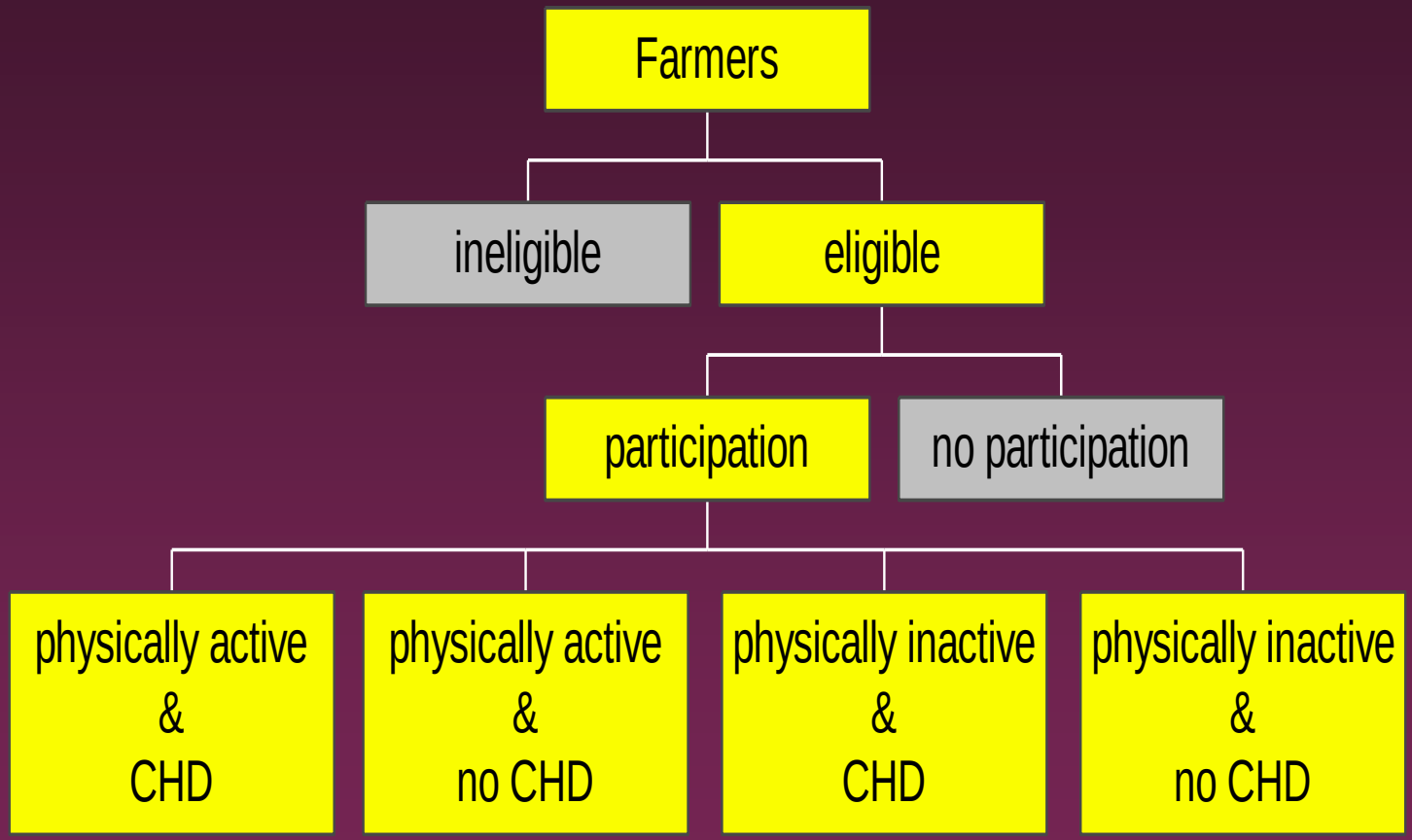
$$\frac{a}{a+b} \text{ vs. } \frac{c}{c+d}$$

OR

Prevalence of exposure compared in diseased and nondiseased

$$\frac{a}{a+c} \text{ vs. } \frac{b}{b+d}$$

Cross-Sectional Study



Cross-Sectional Study

	CHD		
Physically active	yes	no	total
yes	3	87	90
no	14	75	89

Cross-Sectional Study

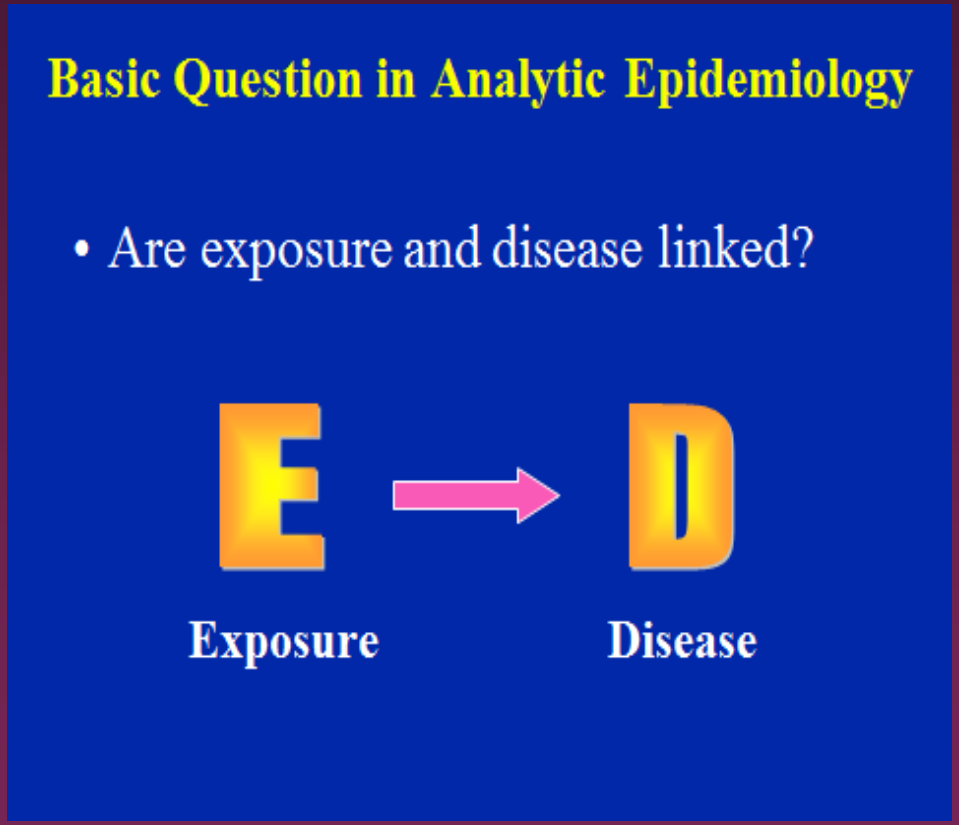
- Advantages
  - Quick and cheap
  - Descriptive role
  - Examine associations
- Disadvantages
  - Temporal associations not clear
  - Shows association, not causality
  - Only a snapshot at a time leading to a misinformation
  - Response rate may be low ,with result not representative of the population

Applications of Descriptive Epidemiological Designs

- Case report, case series, X-sectional,
  - describe disease occurrence
  - natural history of disease
  - generate hypothesis

ANALYTICAL STUDIES

- **Analytic studies**
  - **describe association between exposure and outcome**



Basic Questions in Analytic Epidemiology

- Look to link exposure and disease
  - **What is the exposure?**
  - **Who are the exposed?**
  - **What are the potential health effects?**
  - **What approach will you take to study the relationship between exposure and effect?**

ANALYTICAL STUDIES

Two basic designs:

- Case – control or retrospective study
- Cohort or prospective

NOTE

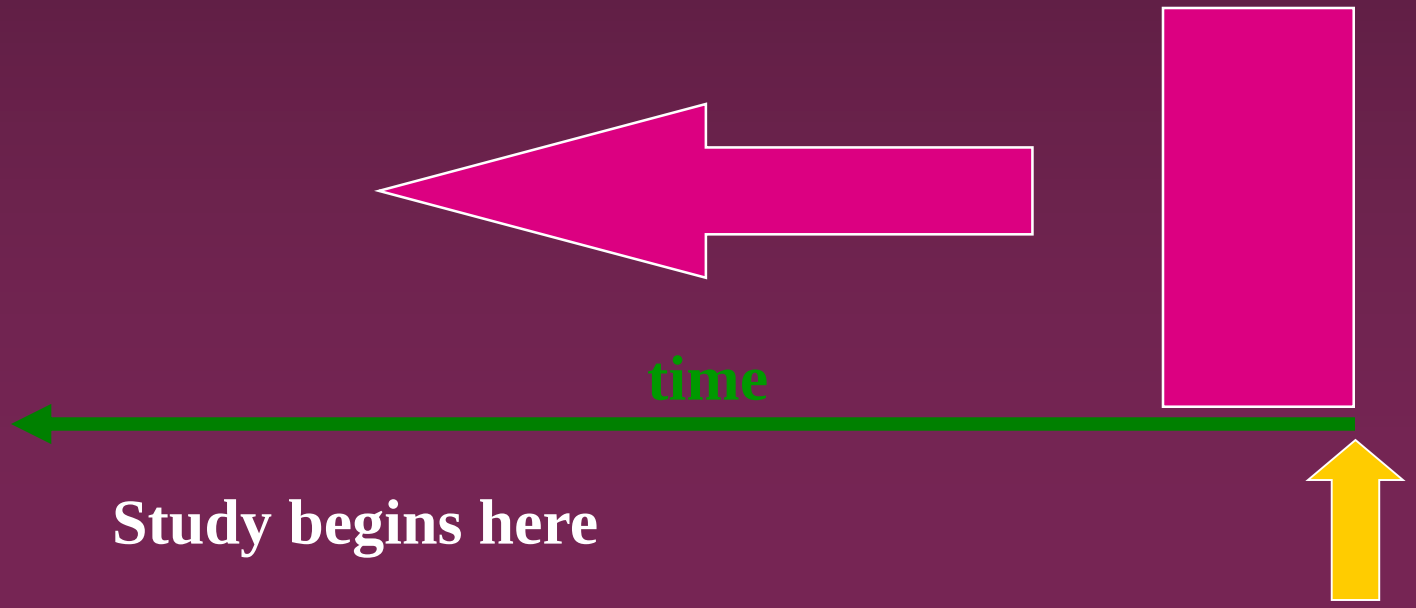
- There must be a comparison group
- No control No conclusion(NCNC)

CASE – CONTROL STUDIES

- an “observational” design comparing exposures in disease cases vs. healthy controls from same population
- exposure data collected retrospectively
- most feasible design where disease outcomes are rare

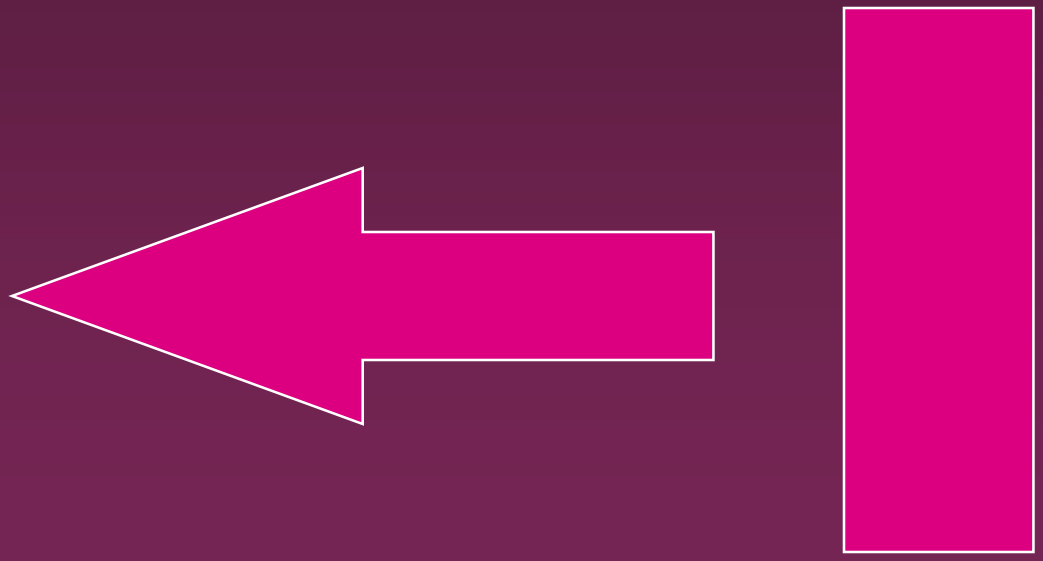
Timeframe of CCS

- **Retrospective Study** - "to look back", looks back in time to study events that have already occurred



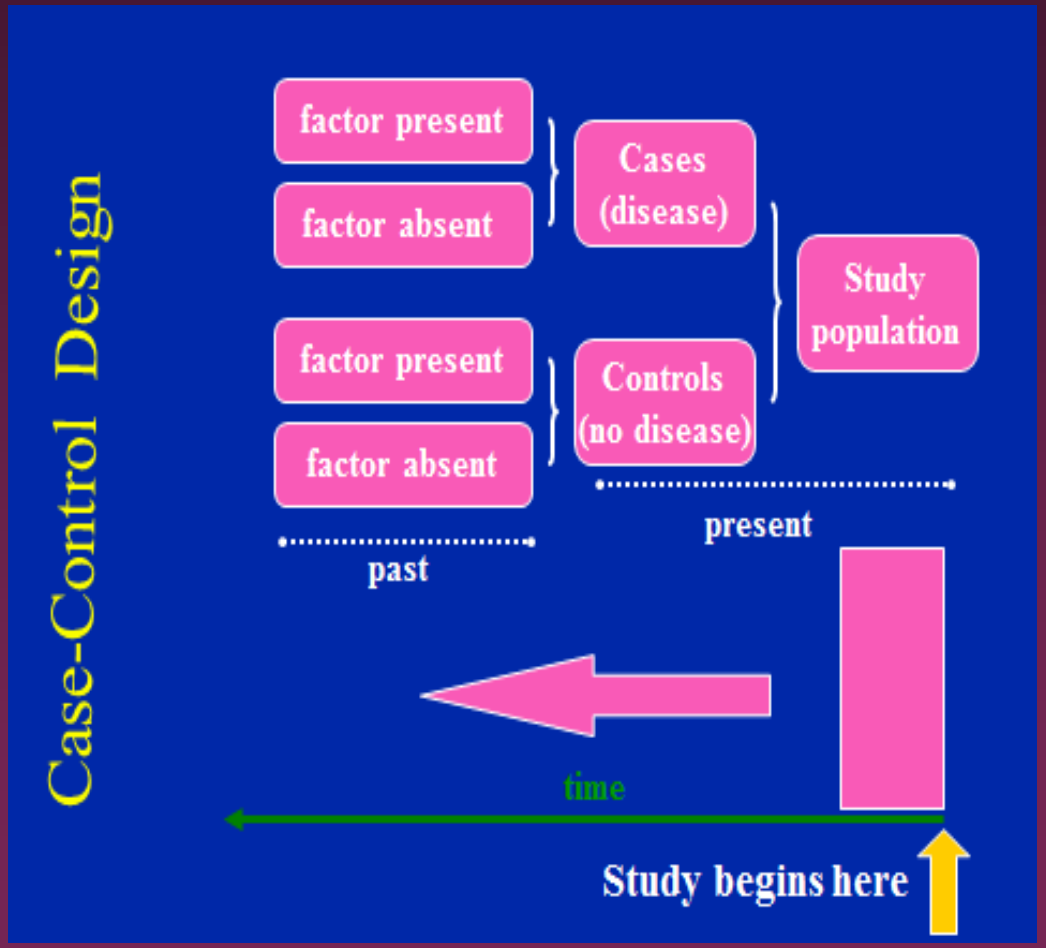
Case-Control Studies

Cases: Disease
Controls: No disease



CASE –CONTROL STUDY

- Question:
  - How do diseased cases differ from non diseased (controls) with respect to prior exposure history?
  - Compare frequency of exposure among cases and controls
  - Effect ⟶ cause.
  - Cannot calculate disease incidence rates because the CCS does not follow a disease free-population over time



Essential features of CCS design

- **Directionality**
  - Outcome to exposure
- **Timing**
  - Retrospective for exposure.
- **Rare or new disease**
  - Design of choice if disease is rare or if a quick "answer" is needed (cohort design not useful)
- **Challenging**
  - The most difficult type of study to design and execute
- **Design options**
  - Population-based vs. hospital-based

Selection of Cases

- Requires case-definition:
  - Need for standard diagnostic criteria e.g., AMI
  - Consider severity of disease? e.g., asthma
  - Consider duration of disease
    - prevalent or incident case?
- Requires eligibility criteria
  - Area of residence, age, gender, etc

Sources of Cases

- Population-based
  - identify and enroll all incident cases from a defined population
  - e.g., disease registry, defined geographical area, vital records
- Hospital-based
  - identify cases where you can find them
    - e.g., hospitals, clinics.
  - But……
    - issue of representativeness?
    - prevalent vs incident cases?

Selection of Controls

- Controls reveal the ‘normal’ or ‘expected’ level of exposure in the population that gave rise to the cases.
- Issue of *comparability* to cases – concept of the “*study base*”
  - Controls should be from the same underlying population or study base that gave rise to the cases?
  - Need to determine if the control had developed disease would he or she be included as a case in the study?
    - If no then do not include
- Controls should have the same eligibility criteria as the cases

Sources of Controls

- Population-based Controls
  - ideal, represents exposure distribution in the general population, e.g.,
    - driver's license lists (16+)
    - Medicare recipients (65+)
    - Tax lists
    - Voting lists
    - Telephone RDD survey

» But if low participation rate = response bias (selection bias)

Sources of Controls

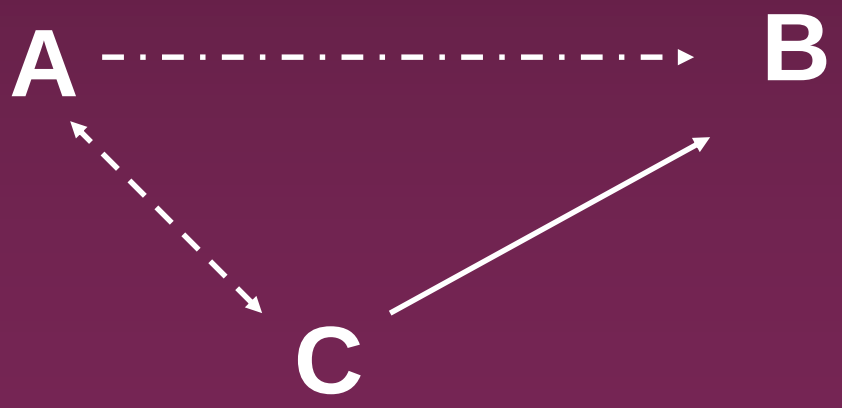
- Hospital-based Controls
  - Hospital-based case control studies used when population-based studies not feasible
  - Advantages
    - more likely to participate (they are sick)
    - efficient (interview in hospital)
  - Disadvantages
    - they have disease?
      - Don't select if risk factor for their disease is similar to the disease under study e.g., COPD and Lung CA
    - are they representative of the study base?

Other Sources of Controls

- Relatives, Neighbors, Friends of Cases
  - Advantages
    - similar to cases SES/ education/ neighborhood
    - more willing to co-operate
  - Disadvantages
    - more time consuming
    - cases may not be willing to give information?
    - may have similar risk factors (e.g., smoke, alcohol)

Controlling extraneous variables (confounding)

- Exposure of interest may be confounded by a factor that is associated with the exposure **and** the disease i.e., is an independent risk factor for the disease



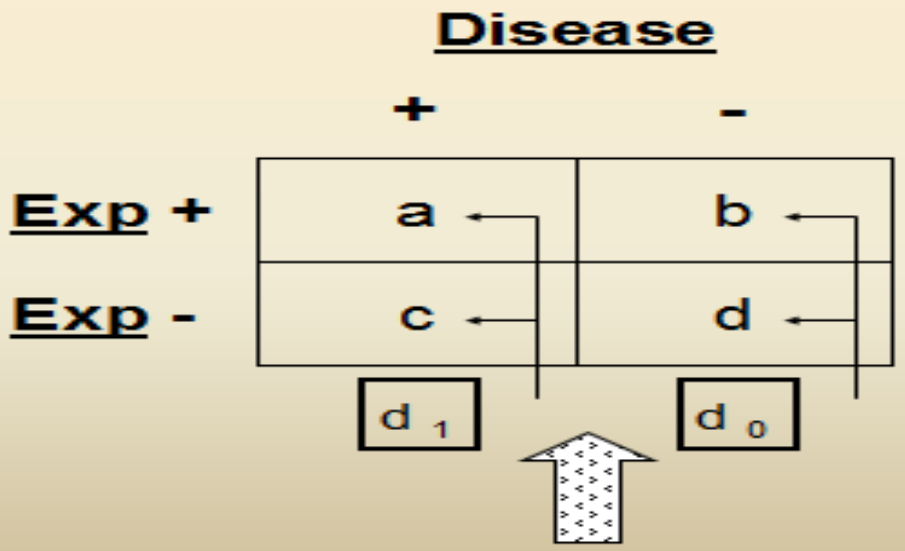
Matching is commonly used in CCS

- Control an extraneous variable by matching controls to cases on a factor you know is an important risk factor or marker for disease
  - Example:
    - Age (within 5 years)
    - Sex
    - Neighbourhood
- Can increase power by matching more than 1 control per case e.g., 4:1
  - Useful if few cases are available

CASE –CONTROL STUDY

Case-control Study – Design

Select subjects on the basis of disease status



8

Odds of exposure among cases = a / c

Odds of exposure among controls = b / d

The OR as a measure of association

- The only valid measure of association for the CCS is the *Odds Ratio* (OR)
- Under reasonable assumptions (the rare disease assumption) the **OR approximates the RR**.
- OR = Odds of exposure among cases (disease) / Odds of exposure among controls (non-dis)
  - Odds of exposure among cases = a / c
  - Odds of exposure among controls = b / d
  - Odds ratio = $\frac{a/c}{b/d} = \frac{a.d}{b.c}$ [= cross-product ratio]

Odds Ratio (OR)

- Similar interpretation as the Relative Risk
- OR = 1.0 (implies equal odds of exposure - no effect)
- ORs provide the exact same information as the RR if:
  - controls represent the target population
  - cases represent all cases
  - rare disease assumption holds (or if case-control study is undertaken with population-based sampling)
- Remember:
  - OR can be calculated for any design but RR can only be calculated in RCT and cohort studies
  - The OR is the only valid measure for CCS
  - Publications will occasionally mis-label OR as RR (or vice versa)

Example - Smoking and Myocardial Infarction

Study: identify all MI cases that occurred over last year (N=40), obtain a random sample of N=40 controls (no MI). What is the association between smoking and MI?

	MI +	MI -
Smk +	30	20
Smk -	10	20
	40	40

$$OR = \frac{a \cdot d}{c \cdot b} = \frac{30 \cdot 20}{10 \cdot 20} = 3.0$$ **(same as the RR!)**

CCS - Advantages

- Quick and cheap (relatively)
  - » so ideal for outbreaks
- Can study rare diseases (or new)
- Can evaluate multiple exposures (fishing trips)

Case-control Studies - Disadvantages

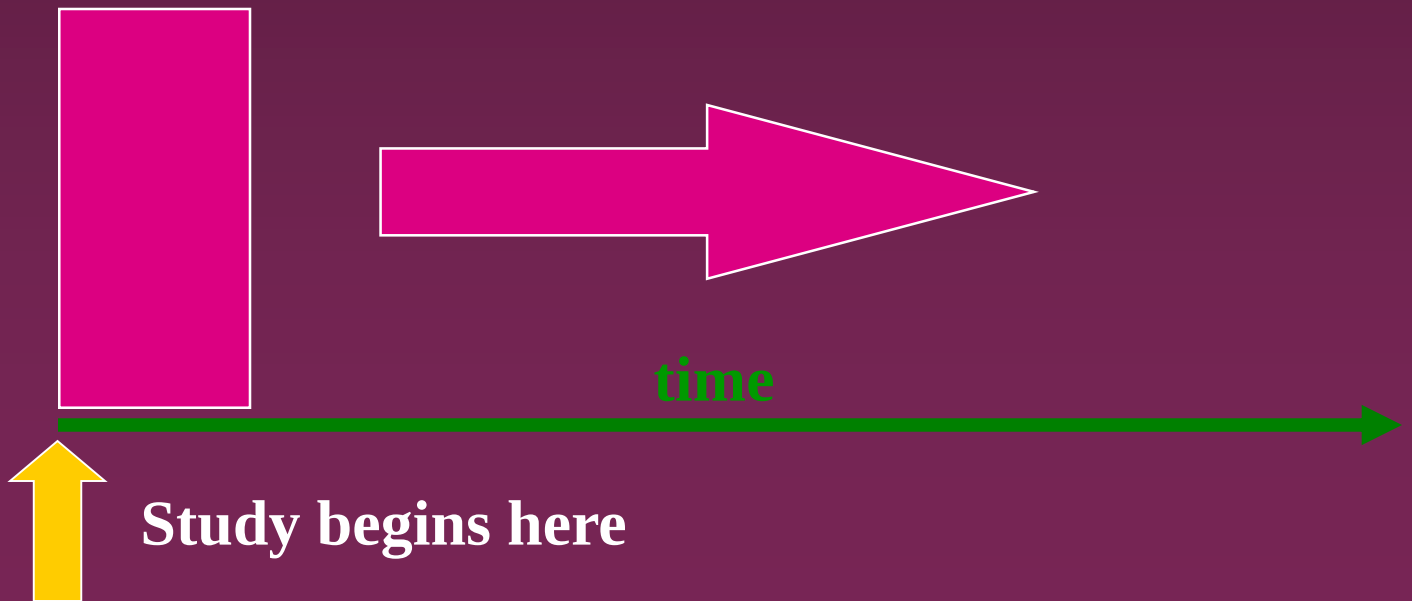
- uncertain of E ⟶ D relationship (esp. timing)
- cannot estimate disease rates
- worry about representativeness of controls
- inefficient if exposures are rare
- Bias:
  - Selection
  - Confounding
  - Measurement (especially recall bias)

Cohort Studies

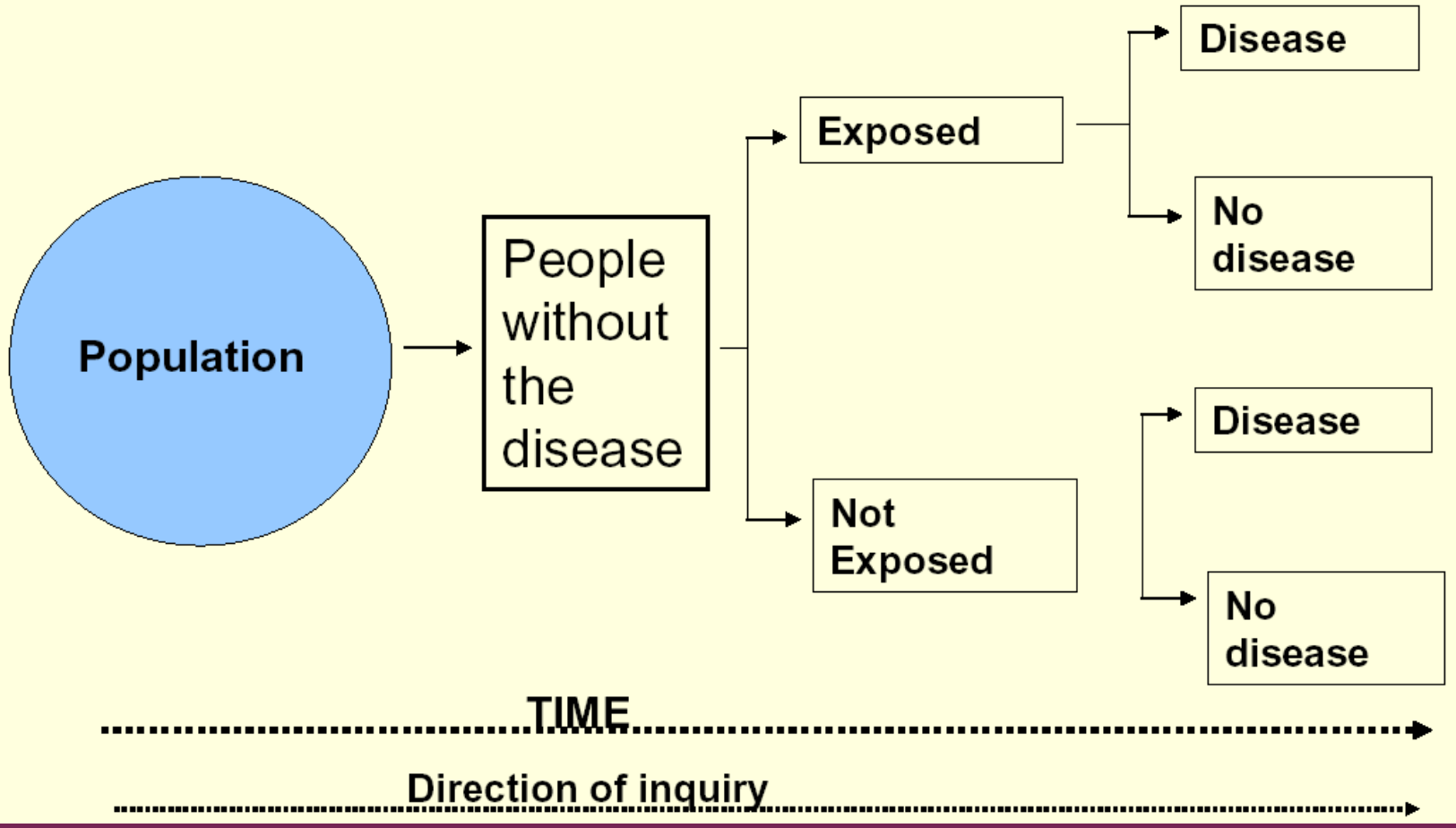
- an “observational” design comparing individuals with a known risk factor or exposure with others without the risk factor or exposure
- looking for a difference in the risk (incidence) of a disease over time
- best observational design
- Also called a longitudinal study or prospective study

Timeframe of Cohort Studies

- **Prospective Study** - looks forward, looks to the future, examines future events, follows a condition, concern or disease into the future

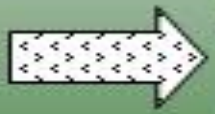


Classical Design of a Cohort Study



Example - Smoking and Myocardial Infarction (MI)

Study: Desert island, pop= 2,000 people, smoking prevalence= 50%
Population-based cohort study. Followed for one year.
What is the risk of MI among smokers compared to non-smokers?



	MI +	MI -	
Smk +	30	970	Rate = 30 / 1000
Smk -	10	990	Rate = 10 / 1000

RR= 3

Relative Risk – Cohort Study

- RR = Incidence rate in exposed / Incidence rate in non-exposed
- The RR is the standard measure of association for cohort studies
- RR describes magnitude and direction of the association :
- RR = 1.0

indicates the rate (risk) of disease among exposed and non-exposed are identical (= null value)

- RR = 2.0

rate (risk) is twice as high in exposed versus non-exposed

- RR = 0.5

rate (risk) in exposed is half that in non-exposed

Example

Presentation of cohort data:
Population at risk

Does HIV infection increase risk of developing TB among a population of drug users?

	Population (follow up 2 years)	Cases
HIV +	215	8
HIV -	289	1

Source: Selwyn et al., New York, 1989

Does HIV infection increase risk of developing TB among drug users?

Exposure	Population (f/u 2 years)	Cases	Incidence (%)	Relative Risk
HIV +	215	8	3.7	11
HIV -	298	1	0.3	

Advantages of Cohort Studies

- Can establish population-based incidence
- Accurate relative risk (risk ratio) estimation
- Can examine rare exposures (asbestos > lung cancer)
- Temporal relationship can be inferred (prospective design)
- Selection and information biases are decreased : Generally less susceptible to bias vs. CCS
- Multiple outcomes can be studied

  (smoking > lung cancer, COPD, larynx cancer)

Disadv. Of Cohort Study

- Expensive and time-consuming
- Inefficient for rare diseases or diseases with long latency
- Large no of subjects are needed
- Unexpected environmental changes may influence the association
- With prolonged time period : Non response, migration and loss-to-follow-up biases

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. Such designs are differentiated 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.

Experimental Studies

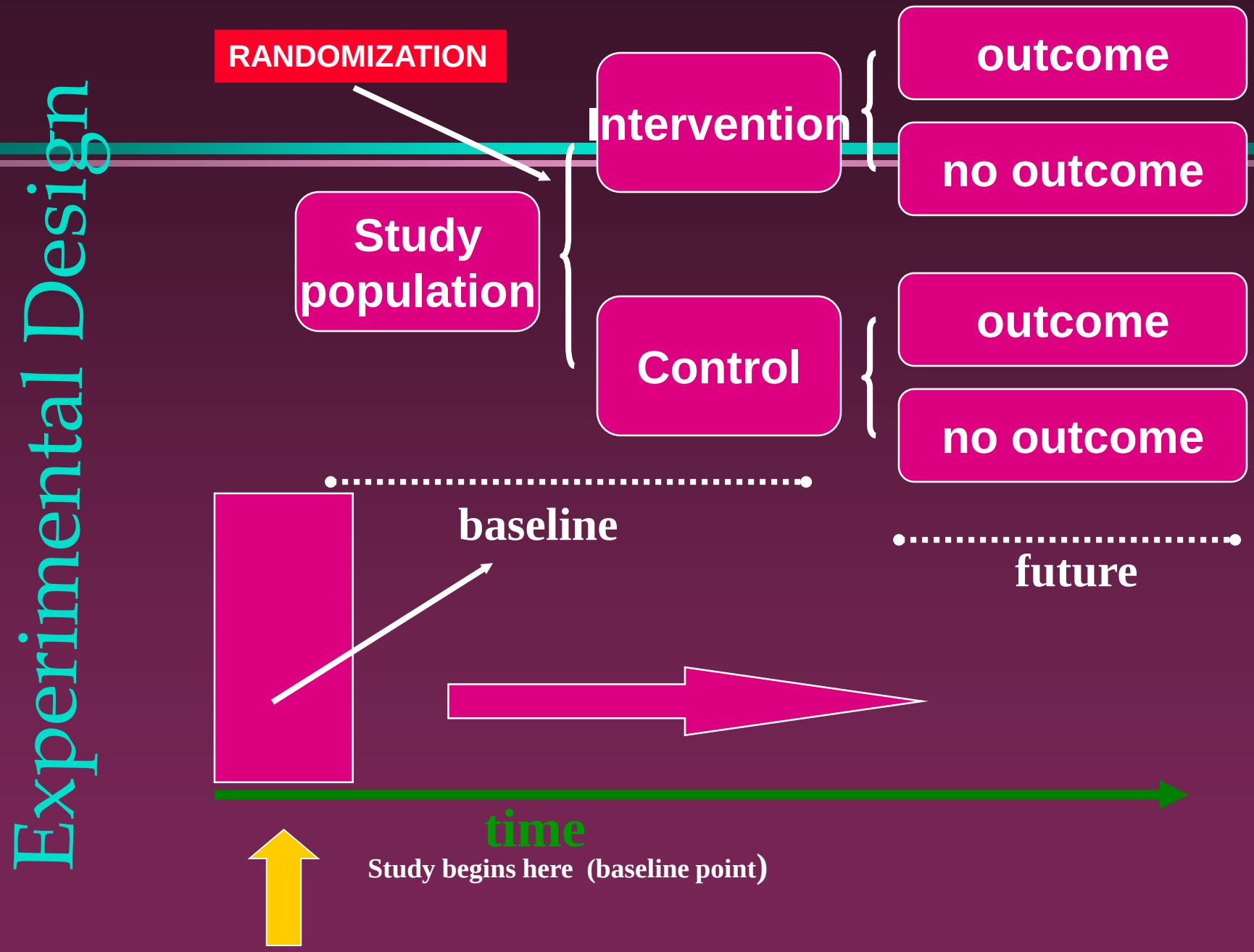
- The investigator can “control” the exposure
- generally involves random assignment to groups
- clinical trials are the most well known experimental design
- the ultimate step in testing causal hypotheses

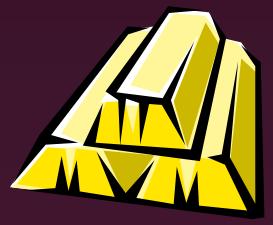
Why Performed ?

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

2. Yield more valid results, as variation is minimized and bias controlled

Experimental Design





RCT Advantages

- **the “gold standard” of research designs. They thus provide the most convincing evidence of relationship between exposure and effect. Example:**
  - ***trials of hormone replacement therapy in menopausal women found no protection for heart disease, contradicting findings of prior observational studies***

- ***Best evidence study design***
- ***No inclusion bias (using blinding)***

RCT Disadvantages

- Long term follow-up (possible losses)
- Compliance
- Expensive
- Possible ethical questions??

How to choose your study design ?

Depends on:

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

Choice of design (II)

It is also related to:

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

Comparing study designs

- Ease
- Timing
- Maintenance and continuity
- Costs
- Ethics
- Data utilisation
- Observer bias
- Selection bias
- Analytic output

Thank You