Descriptive epidemiology

By :YOMIYU(MPH)

Objectives of the session

- Describe the basic concepts of descriptive studies
- Describe the concept of time, place and person in descriptive studies
- Identify the descriptive study
- Describe how to carry out common descriptive studies
- Describe the limitations and strength of Descriptive studies

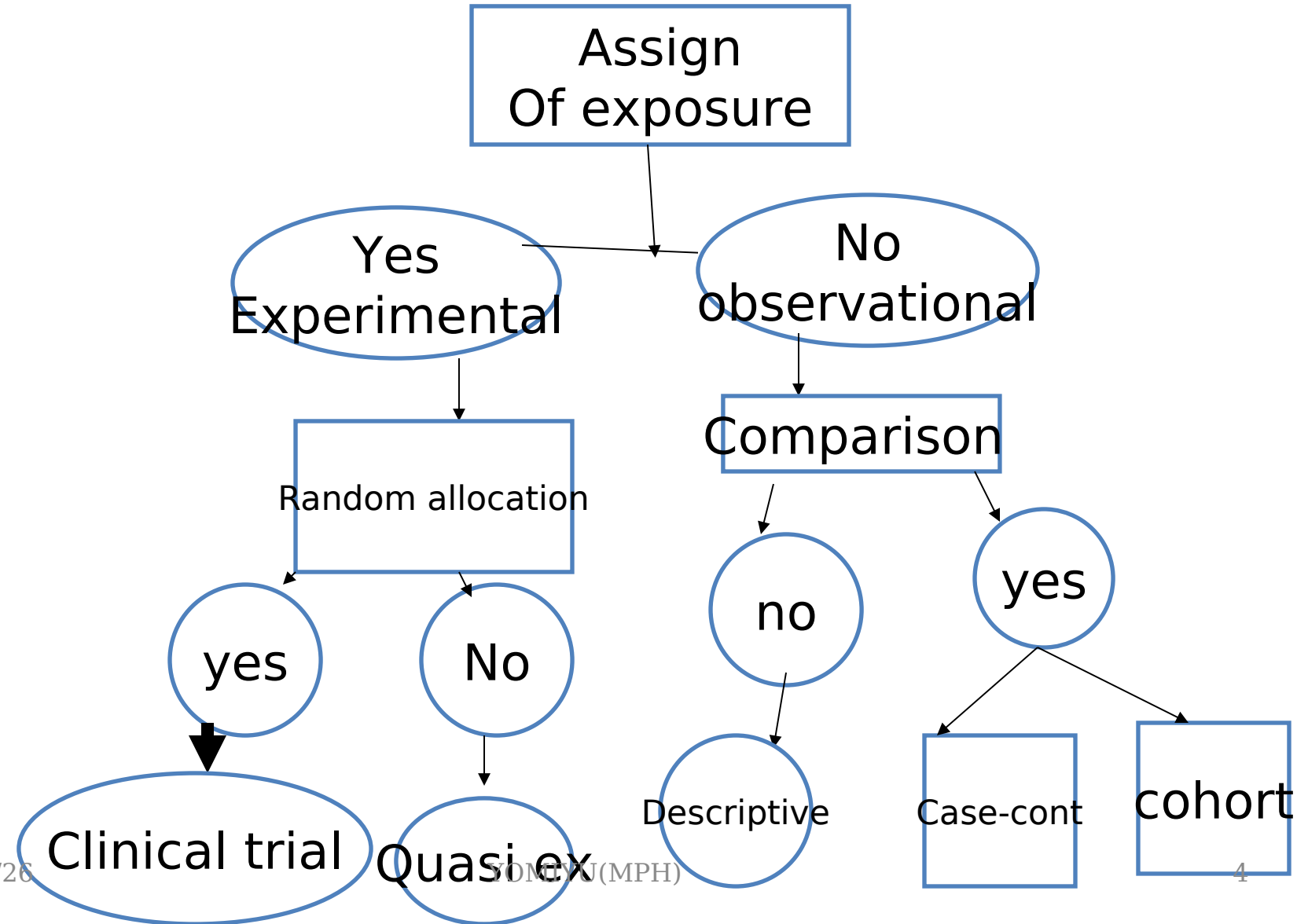
Brain storming Ques..

- What are the types of study designs?

- What is descriptive study?
- What are the types of descriptive study?

- What is the purpose of descriptive study?

Types of study design



DESCRIPTIVE EPIDEMIOLOGY

- Description of disease occurrence, by person, place and time
- Not intended to confirm causes
- **Purpose: To characterize the amount and distribution of disease within a population**

- **In other words …. To identify health problems and patterns of disease that exist.**

Summary on Purpose of Descriptive Epidemiology

- Trend analysis
- Basic to health policy
  - Measuring health
  - Assessing needs
  - Identification of priorities
  - Health planning and evaluation
- Cue/indication about cause

Descriptive…

Descriptive studies generally precede analytic studies designed to investigate determinants of disease.

Thus, descriptive studies often help to generate research hypotheses.

Person

Since disease does not occur at random:

What kinds of people tend to develop a particular disease, and who tends to be spared?
What's unusual about those people?

Person…

<u>Age</u> – the most fundamental factor to consider when describing disease occurrence.

--- The incidence of most chronic diseases increases with age.

--- However, the incidence of many infectious diseases is highest in childhood.

Person…

<u>Gender</u> – biological and non-biological factors related to gender may impact disease risk.

--- In all developed countries, life expectancy is higher in females than males – principally due to lower heart disease mortality.

--- However, many chronic diseases occur more frequently in women (depression, lupus, etc.)

--- As lifestyles continue to become more similar, a question is whether mortality rates will become more similar (i.e. environment vs. biology).

Person…cont

<u>Race/ethnicity</u> – difficult to define, and to identify which characteristics may relate to disease occurrence.

Remarkable variation exists in rates of disease occurrence across racial and ethnic groups.

- **Genetics?**
- **Socioeconomic status?**
- **Environmental exposures?**
- **Access to health care?**
- **Lifestyle factors?**

Person…cont

<u>Social class</u> –Despite its unreliability, SES is consistently associated with mortality in a gradient fashion

Example:

- **Occupation**
- **Education**
- **Area of residence**
- **Income**
- **Lifestyle**

Place

Where is the disease especially common or rare, and what is the different about those places?

Investigation by place includes:

- **Across countries (international)**
- **Within country variation**
- **Urban/rural differences**
- **Localized areas**

Place…

Infectious and chronic diseases show great variation from one country to another.

- **Some differences may be attributed to:**

  --- Climate
  --- Cultural factors
  --- Diet
  --- Genetics

Place...cont

Some differences in disease occurrence between urban and rural locations may be attributed to:

--- Diet
--- Physical activity
--- Housing conditions (i.e. lead paint)
--- Crowding (i.e. spread of infection)
--- Pollution

Time

Since disease does not occur at random:

How does disease frequency change over time?
What other factors are temporally associated with those changes?

The occurrence of health-relate events can vary by time:

- **Secular trends**
- **Cyclic fluctuations**
- **Point epidemics**

Time…cont

<u>Secular trends</u> refer to gradual changes in disease occurrence over long periods of calendar time.

---<u>Example:</u> Death from communicable diseases has been declining where as deaths from chronic diseases has been raising in some developing countries

Time…cont

- **<u>Cyclic fluctuations</u> refer to shorter-term increases and decreases in disease occurrence over a period of years, or within a year.**

Descriptive Study Designs

- **INDIVIDUALS**

  --- Case reports
  --- Case series
  --- Cross sectional surveys

- **POPULATIONS (GROUPS)**

  --- Ecologic (aggregate) studies

Case Reports

- **Careful and detailed report by one or more clinicians of the profile of a single patient**
- **Document unusual medical occurrences**
- **Can provide clues in identification of a new disease or adverse effects of exposures(i.e. a *case report* gave the clue OC use increases the risk of venous thromboembolism)**

Case report...cont

Limitations:

- **No appropriate comparison group**
- **Cannot be used to test for presence of a valid statistical association**
- **Since based on the experience of one person:**
  --- presence of any risk factor maybe purely coincidental
  --- *not a true epidemiologic design*

Case Series (individual data)

- **Description of clinical/epidemiologic characteristics of a number of patients with a given disease**
- **Collection of individual case reports occurring within a fairly short period of time**

Case series…cont

Strengths:

- **Used as an early means to identify the beginning or presence of an epidemic**
- **Can suggest the emergence of a new disease (i.e. AIDS**

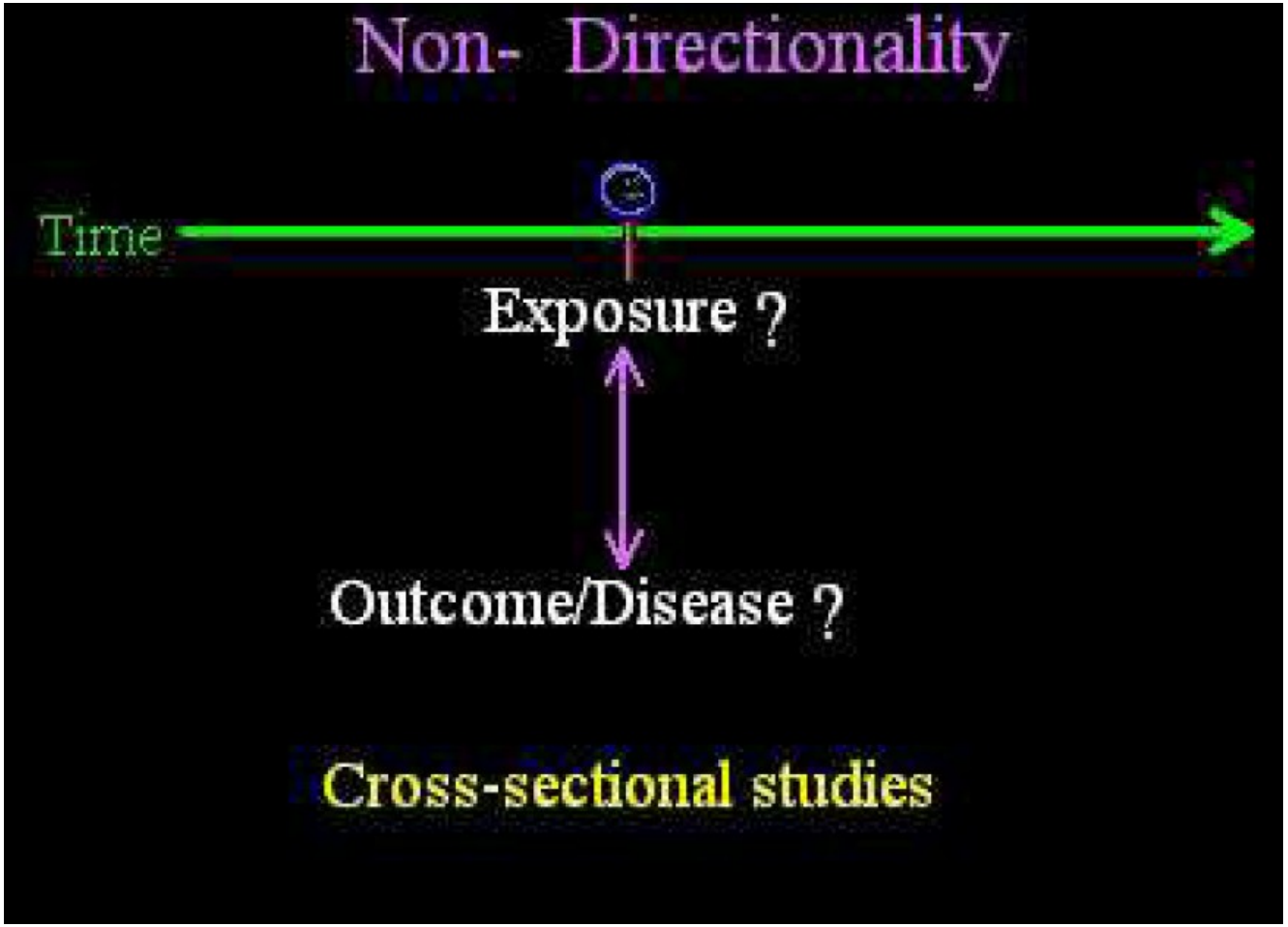
Case series…cont

Limitations:

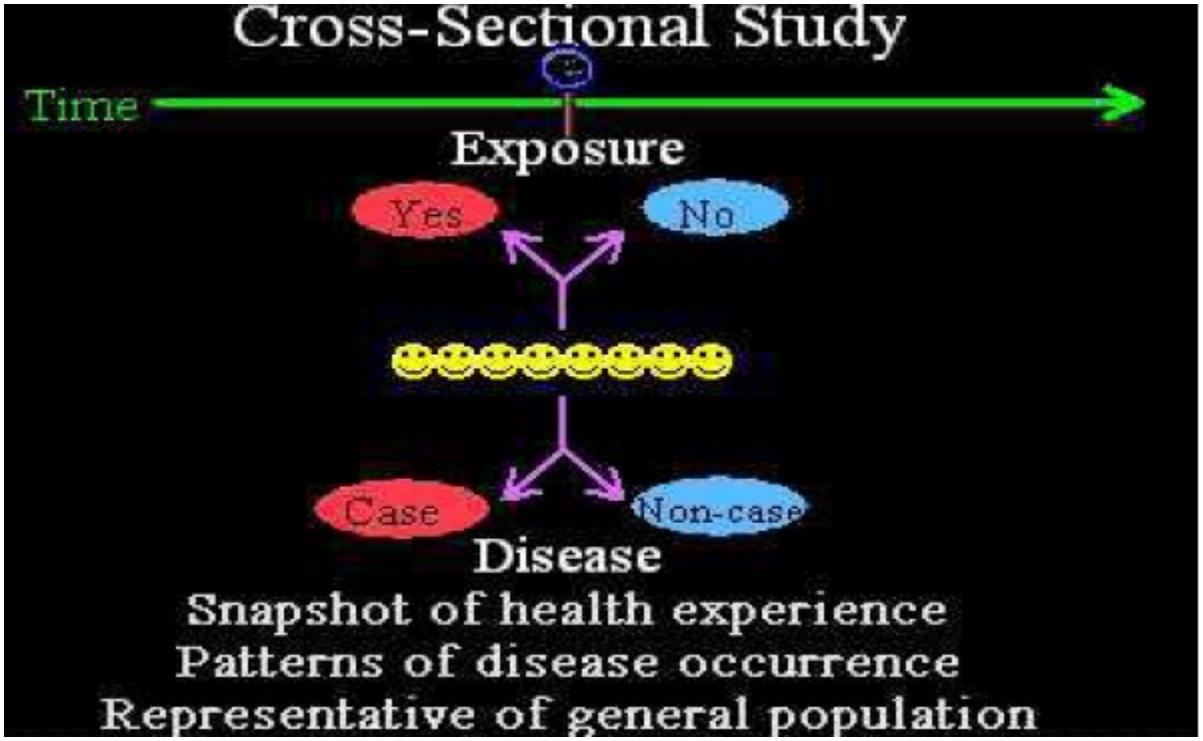
- **Lack of an appropriate comparison group**
- **Cannot be used to test for presence of a valid statistical association**
- ***Not a true epidemiologic design***

Cross-Sectional Studies

- **Snapshot of the health status of populations at a certain point in time**
- **For each subject, exposure and disease outcome are assessed simultaneously (hence also called a "prevalence study/survey")**
- **Compare prevalence of disease in persons with and without the exposure of interest**


Non- Directionality
Time
Exposure ?
Outcome/Disease ?
Cross-sectional studies

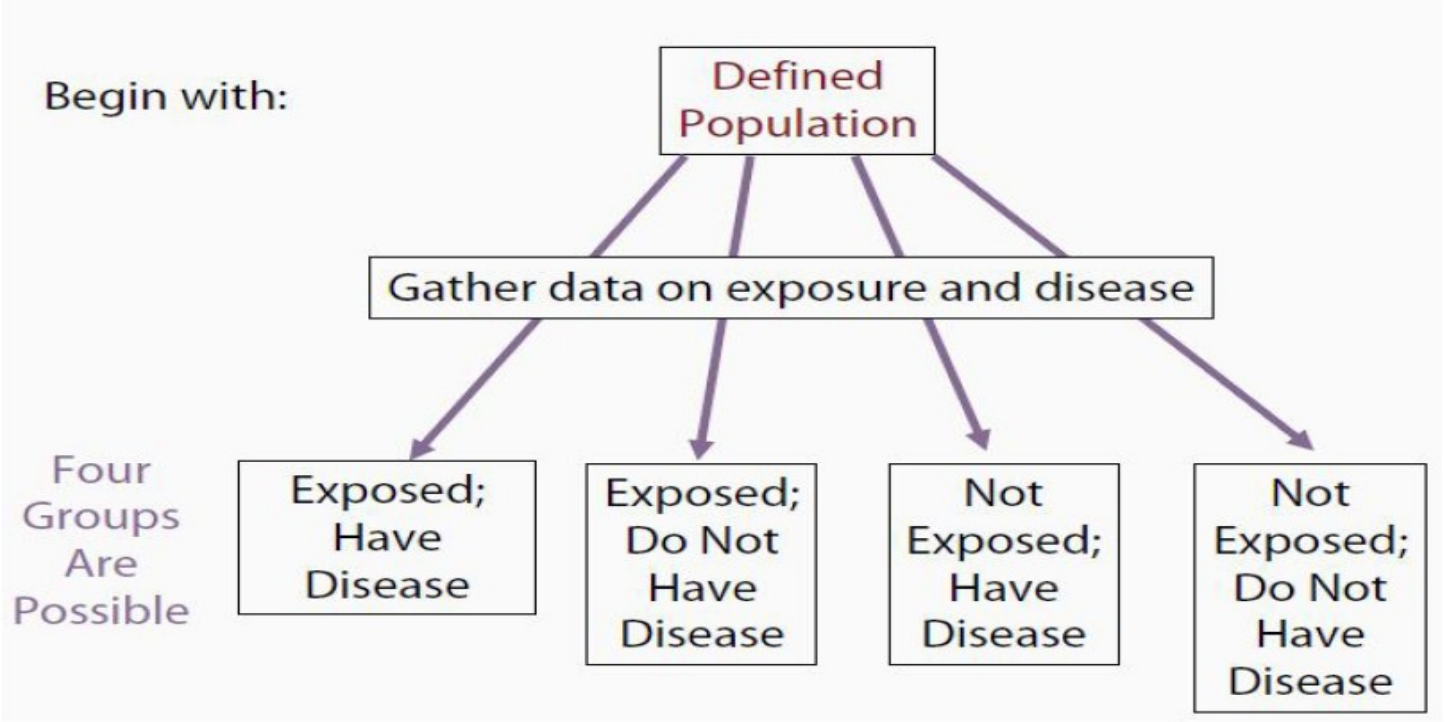
Cross-sectional



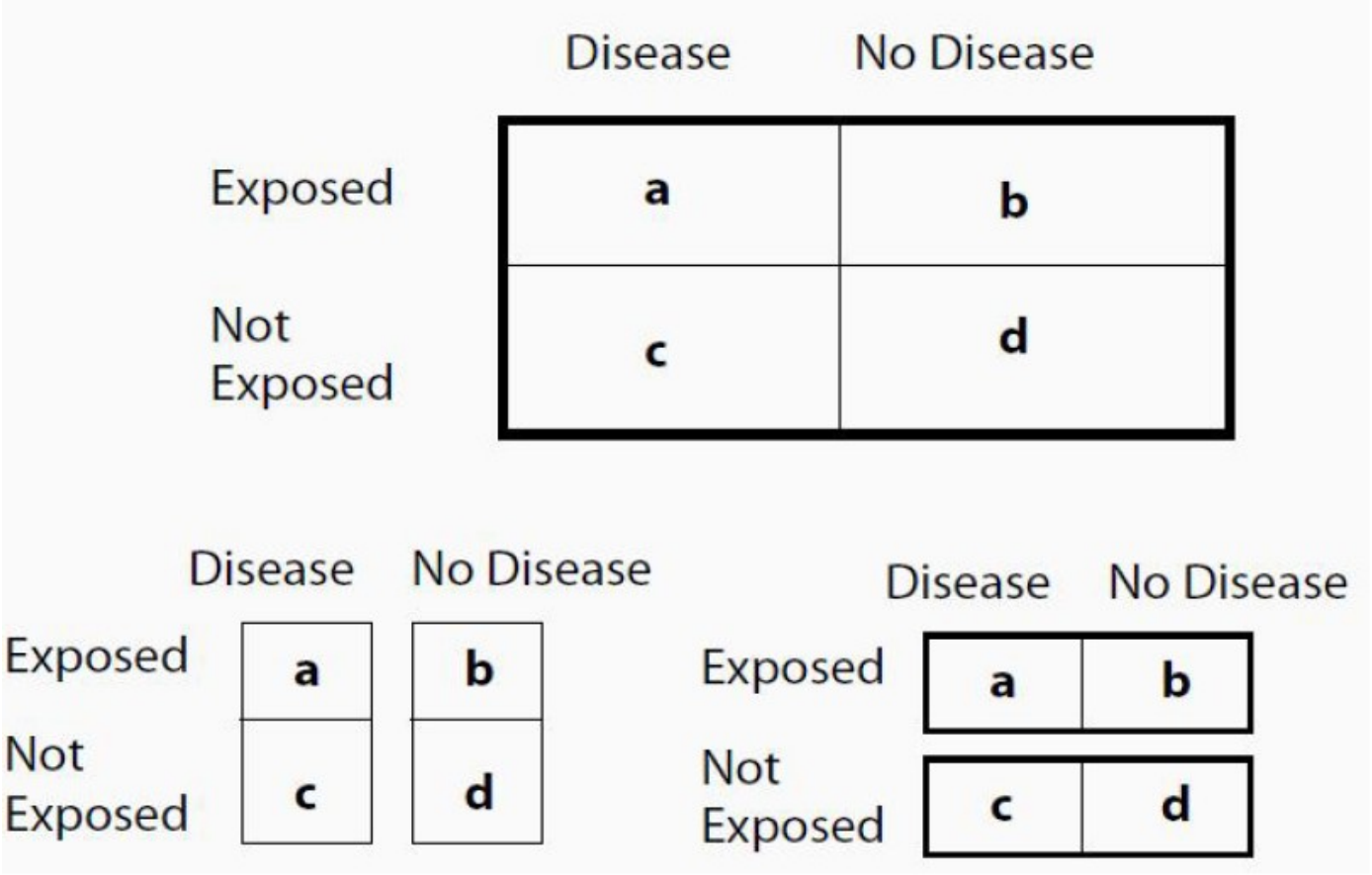
Directionality: Always Non-directional

Timing: Always Retrospective

Design of cross-sectional studies



Design and Analysis of cross -sectional studies



Design and Analysis of cross-sectional studies

	Disease	No Disease
Exposed	a	b
Not Exposed	c	d

Prevalence of exposure in Disease and No Disease

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

	Disease	No Disease
Exposed	a	b
Not Exposed	c	d

Prevalence of disease in Exposed compared to Not Exposed

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

Lack of Temporal Relationship pose difficulty in interpreting findings

Knowledge	Compliance of ART		
	Good	Poor	
Good	950	50	1000
Poor	800	200	1000
	1750	250	2000

OR= 950*200/800*50 = 4.5

Is there association between Know ledge and Compliance?

Which factor is the cause ?

Possibility of establishing temporal relationship in a cross sectional study

Gender	Compliance		
	Good	Poor	
Female	950	50	1000
Male	800	200	1000
	1750	250	2000

OR= 800*50/950*200= 0.21

Is there association between gender and compliance?

Which factor is the cause?

Cross sectional…cont

Strengths:

- **Provides prevalence estimates of exposure and disease for a well-defined population**
- **Easier to perform than studies that require follow-up (hence relatively inexpensive)**
- **Can evaluate multiple risk (and protective) factors and health outcomes at the same point in time**

Cross sectional…

Strengths:

- **May identify groups of persons at high or low risk of disease**
- **Can be used to generate hypotheses about associations between predictive factors and disease outcomes**

Cross sectional…

Limitations:

- **Prevalent rather than incident (new) cases are used – the exposure could be associated with survival after disease occurrence, rather than development of the disease**

- **Temporal sequence between exposure and disease cannot be established**

  *** i.e. Which came first, chicken or the egg?**

Quiz

- A team of researchers from a Black lion hospital decided to estimate the prevalence of diabetic retinopathy among diabetics in AA.
- They chose 1000 consecutive diabetic individuals attending OPD and performed retinal photography on all of them.
- The photographs were divided equally between and read by a panel of 6 Ophthalmologists

- Based on their reports, among those studied, 554 were diagnosed to have diabetic retinopathy. The group concluded that diabetic retinopathy was a major problem among diabetics in AA.

a) What kind of a study design is this?

b) What was the stated objective of the study?

c) What was the population chosen for the study?

d) What are the biases in this study?

e) Is the conclusion of the study valid?

- A team of researchers from a Black lion hospital decided to estimate the prevalence of diabetic retinopathy among diabetics in AA.
- They chose 1000 consecutive diabetic individuals attending OPD and performed retinal photography on all of them.
- The photographs were divided equally between and read by a panel of 6 Ophthalmologists
- Based on their reports, among those studied, 554 were diagnosed to have diabetic retinopathy. The group concluded that diabetic retinopathy was a major problem among diabetics in AA.
- a) What kind of a study design is this?
- b) What was the stated objective of the study?
- c) What was the population chosen for the study?
- d) What are the biases in this study?
- e) Is the conclusion of the study valid?
-

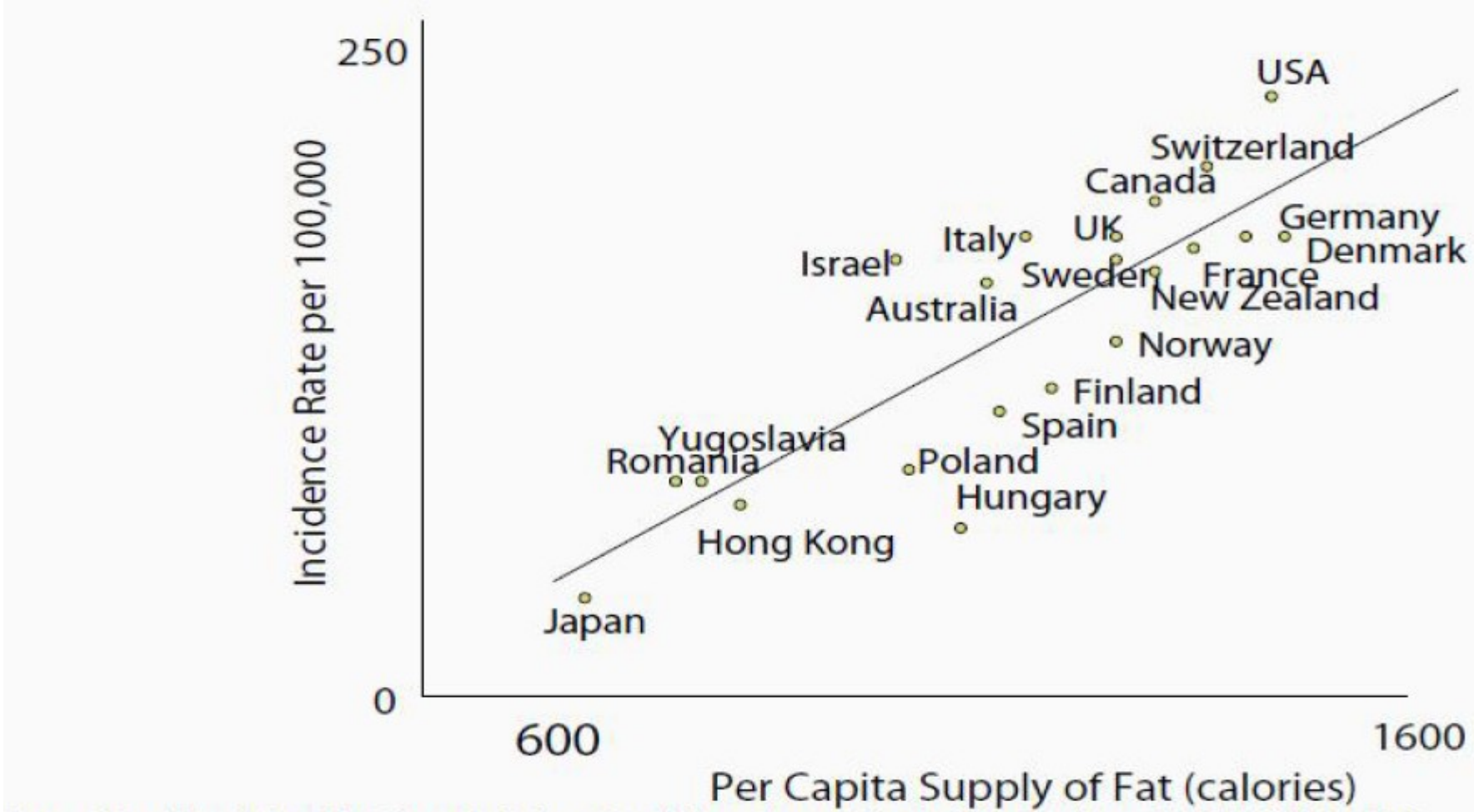
Ecological study

- An **ecological study** is one in which the units of analysis are **populations or groups of people**, rather than individuals
- Example
  - Study of leukemia incidence and exposure to volatile organic chemicals by town
  - Study of prostate cancer mortality and dietary consumption of lycopene in tomatoes by country
- Gives inference on the association between exposure and outcome at the population level (culture, religion, geography, climate, etc.) rather than at an individual level (genes, individual behaviors)

Ecological

- Compare disease frequencies – among different groups during the same period of time, or in the same population at different points in time.
- Does not provide individual data, rather presents average exposure level in the community
- Correlation coefficient (r) is the measure of association

Correlation between Dietary Fat Intake and Breast Cancer by Country



Source: Prentice, Kakar, Hursting, et al., Aspects of the rationale for the Womens' Health Trial. JNCI 80:802-814, 1988.

Ecological studies (Correlatio n)

- Measures represent characteristics of the entire population
- Exposure and risk factors are known only at the group level
- Disease occurrence is at group level
- Ecologic studies are used to generate hypothesis

Ecological studies (Correlatio n)

- Comparison of groups rather than individuals
- Missing individual information
- Low cost
- Interest in ecological effec ts

Levels of measurements

Aggregate measures

- Summaries of observations at individual level
  e.g., proportion of smokers

Environmental measures

- Physiological characteristics

e.g., air pollution levels

Global measures

- GNP, population density

Measuring correlation studies

Correlation coefficient

- Descriptive measure of association in correlation studies.
- The extent to which there is linear relationship between exposure and disease.
- Value of correlation coeffic ient is between +1 and –1

Levels of analysis

Individual level:

E.g. average level of air *pollution is assigned to each individual, and individual age, sex, ethnicity, smoking status are known*

Partially ecological analysis:

E.g. some variables known for individuals (age, sex, air pollution) but others for the population

(% smokers, % vegetarians)

Fully ecological analysis:

Information on exposure and di sease are only known for the population

Ecological Fallacy

- Across level inference.

  E.g. from group to individual.

The ecologically fallacy

- Concluding that because an association exists between exposure and disease at the group level it therefore exits at the individual level.
- We don't know the link between exposure and disease among individuals with in each group

Ecological fallacy: example -2

- Imagine a study of the rate of coronary heart disease in the capital cities of the world relating the rate to average income. Within the cities studied, coronary heart disease is higher in the richer cities than in the poorer ones.

  We might predict from such a finding that being rich increases your risk of heart disease.

- *In the industrialized world the opposite is the case -within cities such as London, Washington and Stockholm, poor people have higher CHD rates than rich ones.*

The ecological fallacy is usually interpreted as a major weakness of ecological analyses.

- Ecological analyses, however, inform us about forces which act on whole populations.

Types of Ecologic Study Designs

- Cross-sectional ecologic studies
- Time-trend ecologic studies
- Descriptive ecologic studies

1. Cross-sectional Ecologic Studies

- compare aggregate exposures and outcomes across communities over the same time

2. Time-trend Ecologic Studies

- compare variations in aggregate exposures and outcomes over time within the same community.

3. Descriptive Ecologic Studies

- Investigate disease or risk fa ctor differences between communities at the same time, or within the same community over time.

Ecologic Study Designs at a Glance		
Study type	**Design**	**Time frame**
Cross-sectional	Across communities	Same time period
Time-trend	Within the same community	Over time
Descriptive	Across communities or Within the same community	At a point in time or Over time

Advantages of Ecologic Studies

- Aggregate data on exposure and disease are often publicly available
- Aggregate level data can conveniently be obtained by researchers at a low cost
- Can be useful for evaluating the impact of community-level interventions.

Advantages of Ecologic Studies

- In addition, minimal within-community differences between exposures may exist; however, exposures may differ substantially between communities, cities, states, and countries.
- Examples of small within-community exposure differences but large between-community differences include:

– Quality of drinking water

– Concentration of certain air p ollutants such as ozone and fine particles

– Average fat content of diet

– Cumulative exposure to sunligh t

Limitations of ecological studies

1. Inability to link exposure with outcome at individual level
2. Lack of ability to Control potential confounders
3. It represent average exposu re levels rather than actual values
4. Mask non-linear relationship between exposure and disease
5. Ecological Fallacy

Analytical study design
Case control study

Objective of the Session

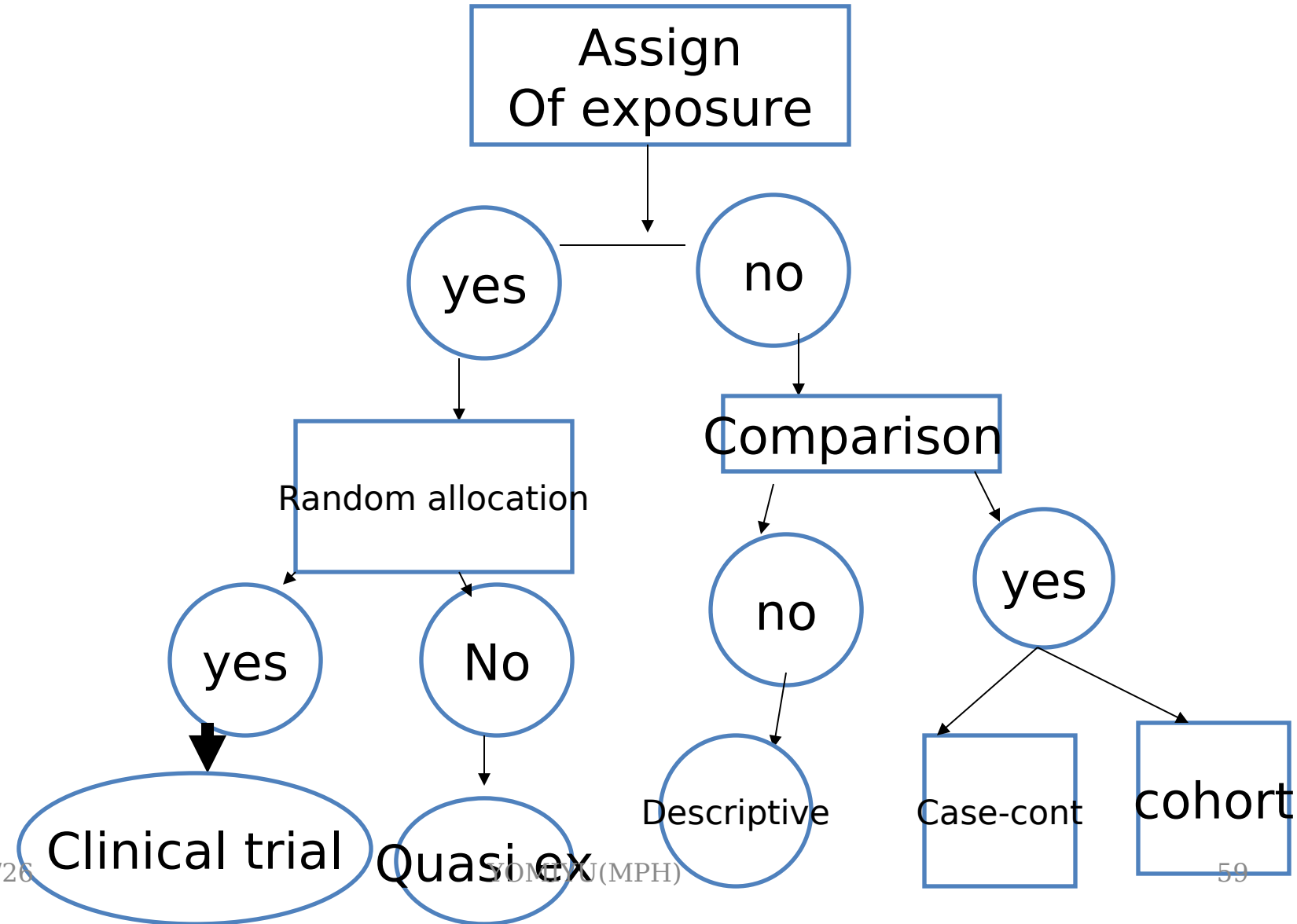
- Describe the basic concepts of case control study
- Describe how to conduct and analyze case control study
- Demonstrate the basic concepts of case control study
- Describe the limitations and strength of case control study

Discussion

- What are the purpose of analytical study designs?
- List types of analytical study designs?
- How do you understand case-control study?
- What type of analytical epidemiology study design do you know?
- What is the purpose of analytical epidemiology?

Analytical vs. Descriptive study

- **Descriptive studies often focus on a single variable such as a particular disease or health state**
- **Unlike descriptive studies, analytic studies in epidemiology must have a clear hypothesis (usually an etiologic hypothesis).**
  - **At least two variables must be specified in advance, an exposure and a disease or health state**

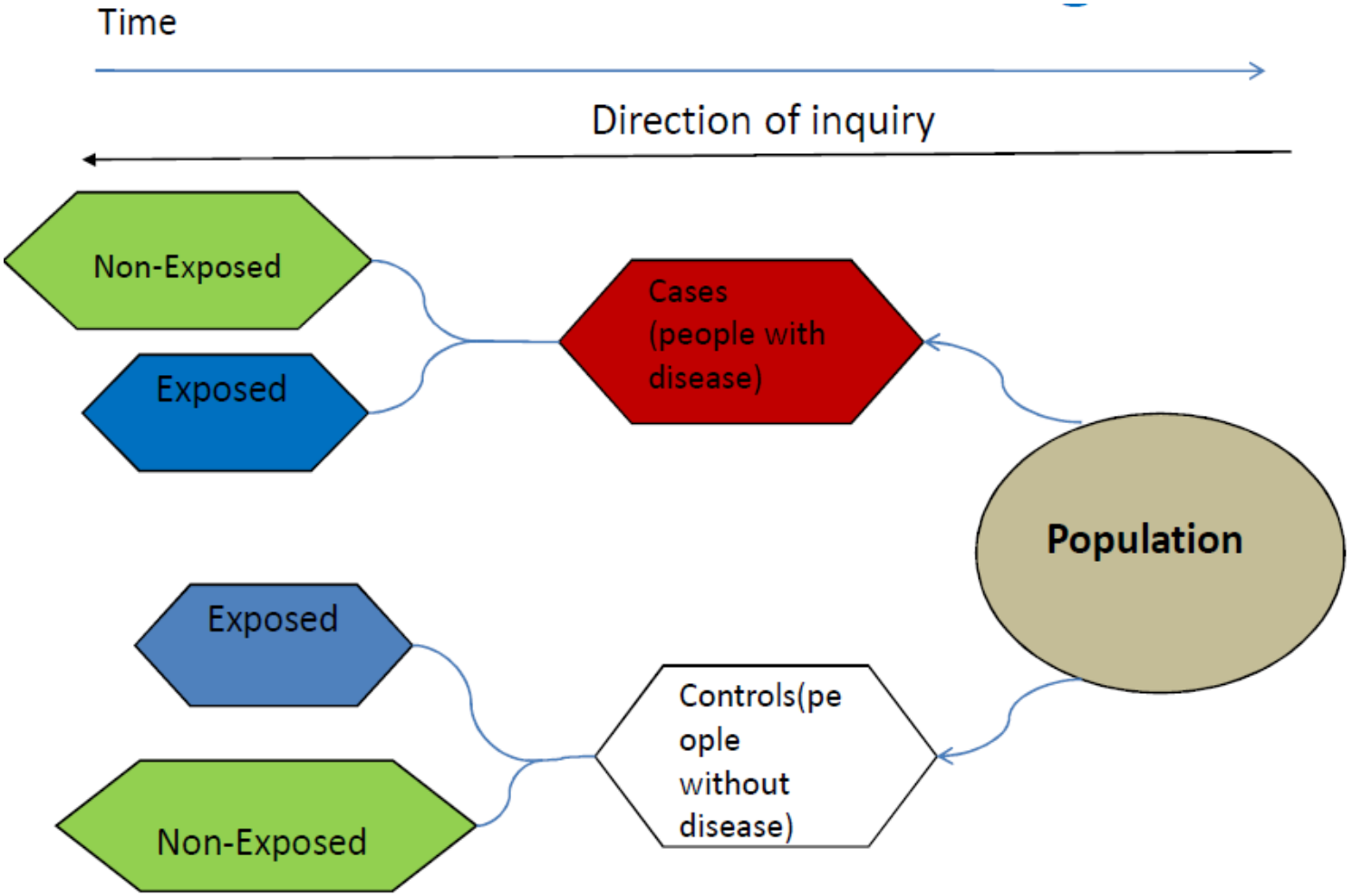

Assign
Of exposure
yes
no
Random allocation
Comparison
yes
No
no
yes
Clinical trial
Quasi ex
Descriptive
Case-cont
cohort

Case control study

I. Design concept

- Starts with cases and comparative groups(controls)
- We determine what proportion of cases were exposed and what proportion were not
- We also determine what proportion of controls were exposed and what proportion were not

Case-control Design



II. Selection of cases

A/ Definition- precise definition of cases(clinical ,laboratory and other criteria

B/ Inclusion and exclusion criteria

-cases should be selected to improve validity(ex. By excluding cases with coexisting disease)

-cases should be restricted to limited time period and geographic area, age range etc…

Selection of cases…cont

C/ Incident or prevalent cases?

To the extent possible avoid prevalent cases.

Why?

- Chicken-egg dilemma on cause and effect
- Non-representative cases since long-time survivors are selected
- Prevalent cases with long disease duration may not accurately recall antecedent events
- Difficult to distinguish prognostic factor and cause

Selection of cases…cont

- **However, case-control studies of congenital malformations are inevitably based on prevalent cases**

Prevalent cases are commonly used in studies of chronic conditions with ill-defined onset times (e.g. multiple sclerosis)

Source of cases

- **Sources of Cases:**
  - **Hospitals, other medical care facilities, etc.**
  - **General population - locate and obtain data from all or a random sample of individuals from a defined population.**

Selection of cases…cont

- **<u>Note:</u> DO NOT compromise validity in the goal of generalization.**

- **Select cases from a defined population in whom <u>complete and reliable information</u> can be obtained, and where the exposure/disease relationship is presumed to be present.**

III. Selection of controls

- **Selection of an appropriate comparison group is the most difficult and critical issue in the design of case-control studies.**
  - Controls are free of the diseases
  - Controls should represent the population at risk of becoming cases(controls and cases should come from the same source population)

Selection of controls…cont

- **The prevalence of exposure among controls should reflect the prevalence of exposure in the source population.**
- **The time during which a subject is eligible to be a control should be the time in which the individual is also eligible to be a case.**

If the above three points are not fulfilled=selection bias

Selecting Controls

Sources of Controls:

--- General population
-- Neighborhood
--- Friends/relatives
--- Hospital or clinic-based

Selection of Controls…cont

General Population Controls:

---Population defined by geographic boundaries (or specific characteristics).

---Cases may include all cases, or a random sample of all cases.

---Controls should be a random sample of non-diseased individuals eligible to be controls.

Selection of control…cont

General Population Controls:

--- If entire population is sampled for cases and controls, can calculate incidence rates of disease in exposed and non-exposed.

--- Selection of controls may be costly, time consuming, and exposure recall may not be as accurate as "sick" controls.

--- Subjects in general population may be less motivated to participate than hospital-based controls.

Controls…cont

Neighborhood Controls:

--- May approximate random sampling from the source population.

--- Controls are often matched to cases from the same neighborhood.

--- If cases are from a particular hospital, neighborhood controls may include people who would have been treated at the same hospital had they developed the disease

Controls…cont

Friend/Relative Controls:

- **Tend to be more cooperative than general population controls.**
- **Often similar to cases on factors such as SES, lifestyle, and ethnic background.**
- **However, being named as a friend by the case may be related to exposure status of the potential control.**

Controls…cont

Friend/Relative Controls:

- **The list of potential friend/relative controls is often derived from the case; this dependence may add a potential source of bias.**

- **Hence, friend/relative controls may be too similar to cases regarding the exposure of interest.**

Controls….cont

Hospital/Clinic-Based Controls:

--Source population refers to people who “feed” the hospital or clinic.

--Usually easier and less expensive than general population controls.

--May be more aware of exposures and likely to cooperate than general population controls (healthier).

Controls…cont

Hospital/Clinic-based Controls:

- **Controls are ill; distribution of the exposure may not reflect the distribution of exposure in the source population for cases.**
- **Controls should be limited to diagnoses for which there is no prior indication of a relation with exposure.**
- **Subjects may have changed their exposure status as a result of being sick.**

Controls…cont

For each control group, how many controls per case?

-- the optimal case-control ratio is 1:1

-- when the number of cases is small, the sample size for the study can be increased by using more than one control

e.g. 1:2 1:3 1:4

Controls…cont

- **The benefit of increased sample size is not as relevant past the 1:4 ratio (e.g increase in statistical power).**

Case-control studies

Advantages

- Useful for studying several potential exposures.
- Efficient for rare diseases.
- Efficient in resources and time
- Relatively quick, disease and exposure measurements can be made at the same point in time.

Disadvantages

- Prone to selection bias and recall b ias.
- Not suitable for rare exposures.
- May be difficult to establish that "cause”preceded "effect”
- Cannot yield a population-level measure of disease incidence or prevalence.

Measure of association in case controls study

- **The measure of association in case control study is Odds Ratio(OR)**

$$= \frac{\text{odds of disease in exposed}}{\text{odds of disease in unexposed}}$$

OR

Odds = $\frac{\text{the chances of something happening}}{\text{the chances of it } \textit{not} \text{ happening}}$

Measure of association... cont

OR= ad/bc

OR=1=Null hypothesis

OR>1 Exposure is a risk factor

OR<1 Exposure is protective

Calculating the Odds Ratio

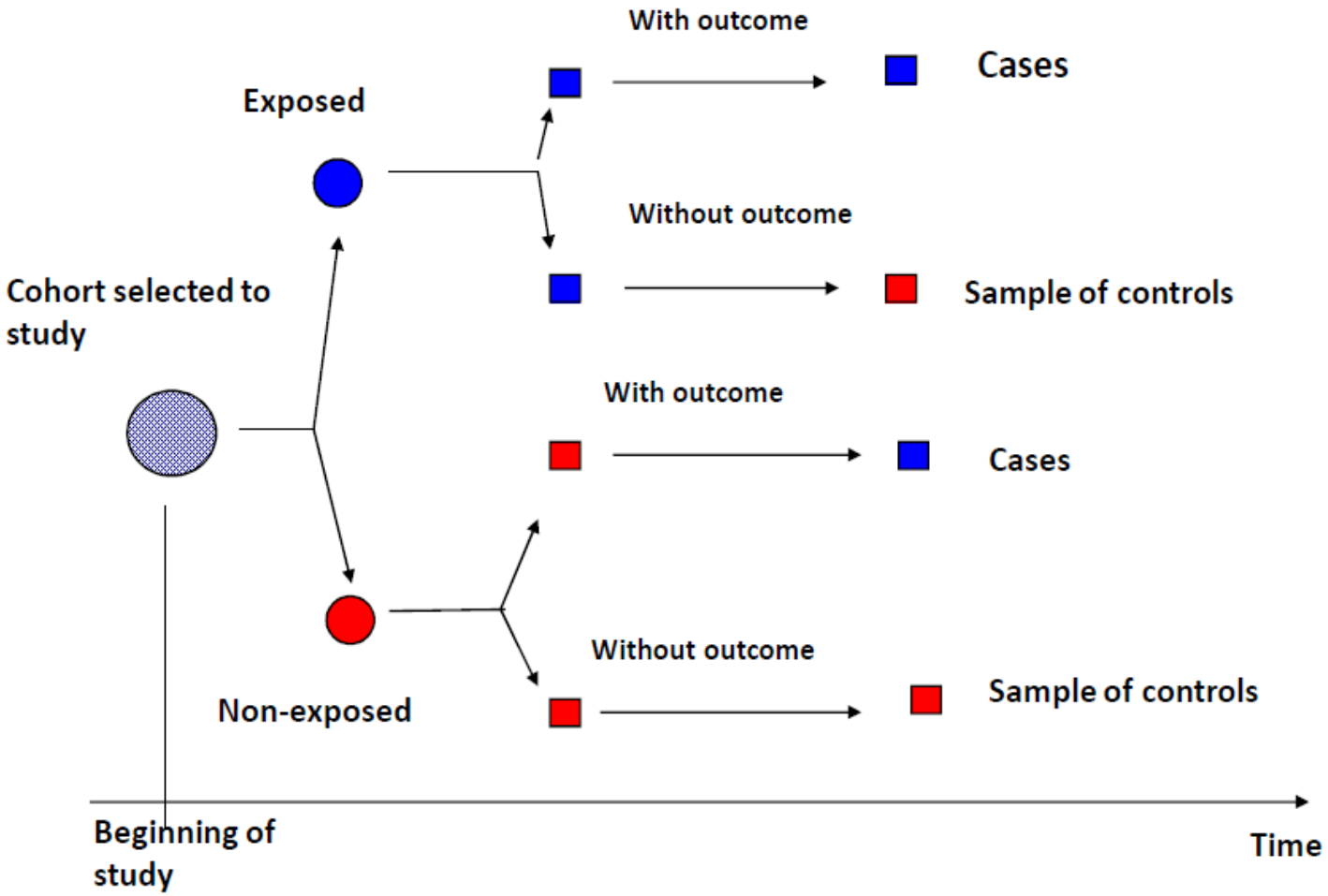
Exposure Status	Disease Status: CHD cases (Cases)	Disease Status: No CHD (Controls)
Smoker	112	176
Non-smoker	88	224
Total	200	400

$$\text{Odds Ratio} = \frac{AD}{BC} = \frac{112 \times 224}{176 \times 88} = 1.62$$

Nested Case Control Studies

- Begin as cohort study, classify the subjects as exposed or non-exposed; in the follow up, some participants develop the outcome, and they are the **cases** and a sample of subjects without outcome are **control**s and with them analyzed the **case-control** study.

Nested Case Control Studies



Nested Case Control Studies

Advantages

- Cases and controls are members of the same population. Hence, the probability of **selection bias** is minimized.

- Information about the exposure is collected before the develop o f outcome, and the probability of **information bias** is minimized.

Nested Case Control Studies

- Is a case-control study conducted within cohorts of exposed and unexposed people.

- The reason to conduct a case-control study even when a cohort can be enumerated is usually that more information is needed than is readily available from records and it would be too expensive to seek this information for every one in the cohort.

COHORT STUDIES

Session objectives and contents

- Objective
  - Describe and apply the basic concepts in the conduct of cohort studies
- Contents
    - Definition and types
    - Design, data collection and data analysis
    - Advantages and Disadvantages
    - Bias and confounding

Definition (1)

- Cohort
  - a group of people who share a common experience or condition
  - E.g. Birth cohorts, cohort of smokers, occupational exposures
- Cohort studies
  - The observation of a cohort over time to measure outcome(s)
  - Because the data on exposure and disease refer to different points in time, cohort studies are longitudinal, like case-control studies
  - Longitudinal, follow-up or incidence studies

Definition (2)

- They have 2 primary purposes:
  - Descriptive: to describe the incidence rates of an outcome
  - Analytic: to analyze associations between the outcomes and risk factors
    - Usual type

Definition (3)

- begin with a group of people free of disease who are classified into subgroups according to exposure to a potential cause of outcome
- and the whole cohort is followed up to see how the subsequent development of outcome differs between the groups with and without exposure (Figure 1)

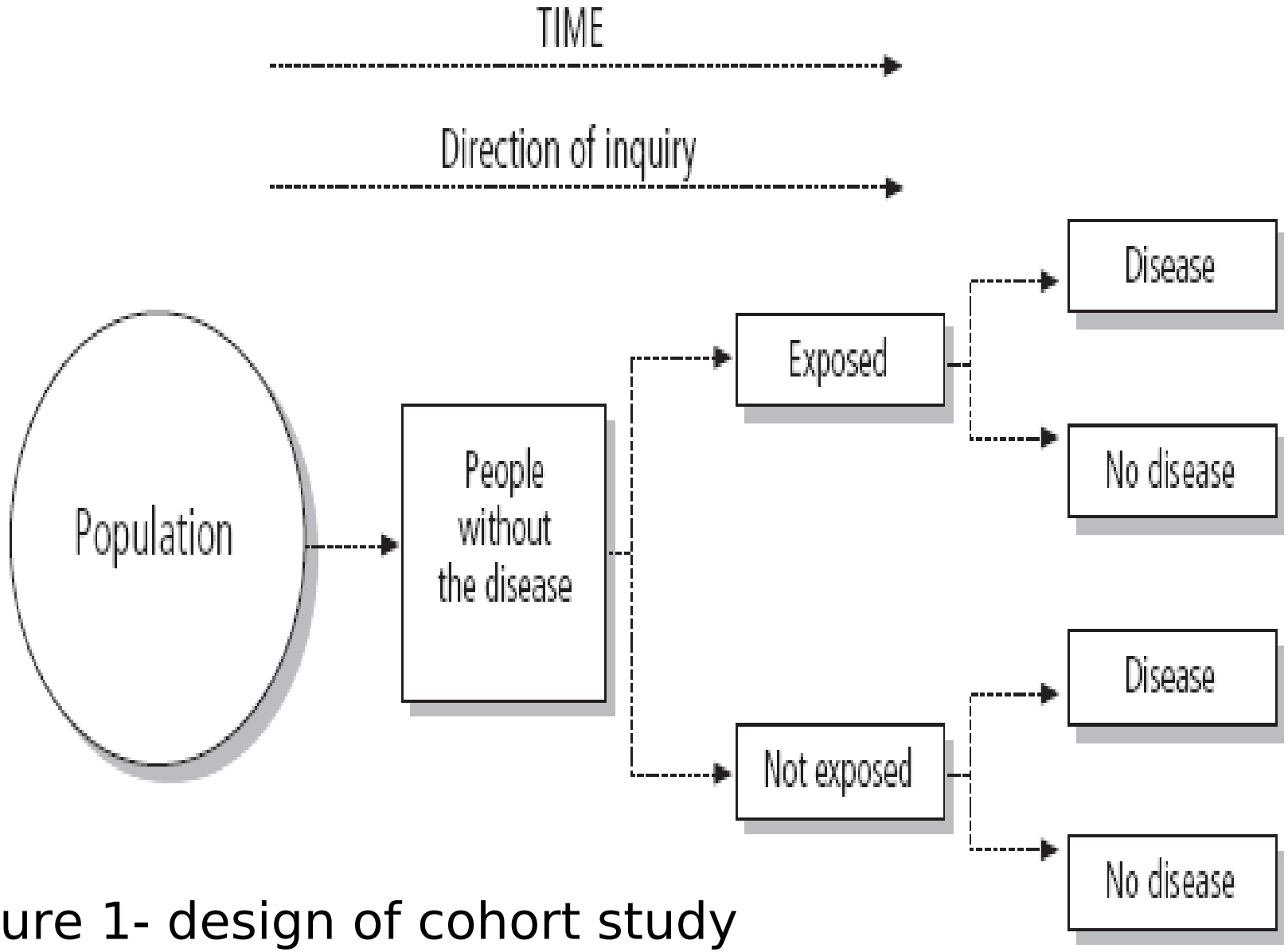


Figure 1- design of cohort study

Types (1)

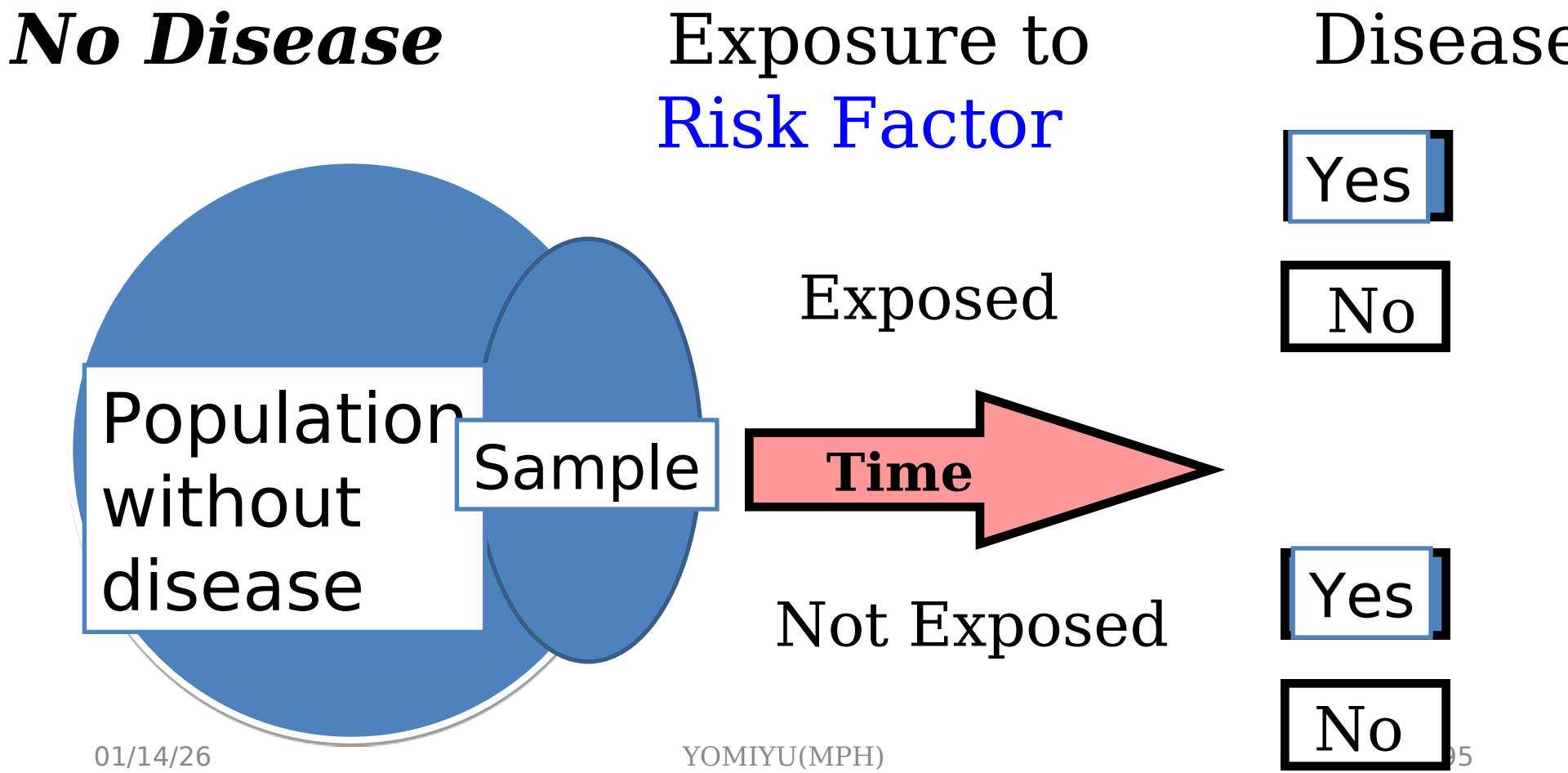
- Closed vs. Open
  - Closed cohort: exposure groups are defined at the start of follow-up and no new members are added during the follow-up
  - Open/dynamic cohort: people move in and out the study

Types (2)

- Incidence cohort vs. Prognostic (clinical)
  - Incidence Cohort Study
    - To assess incidence of disease
    - To identify risk factors for disease onset
    - Incidence greater in exposed than non-exposed?

Incidence cohort

- fig

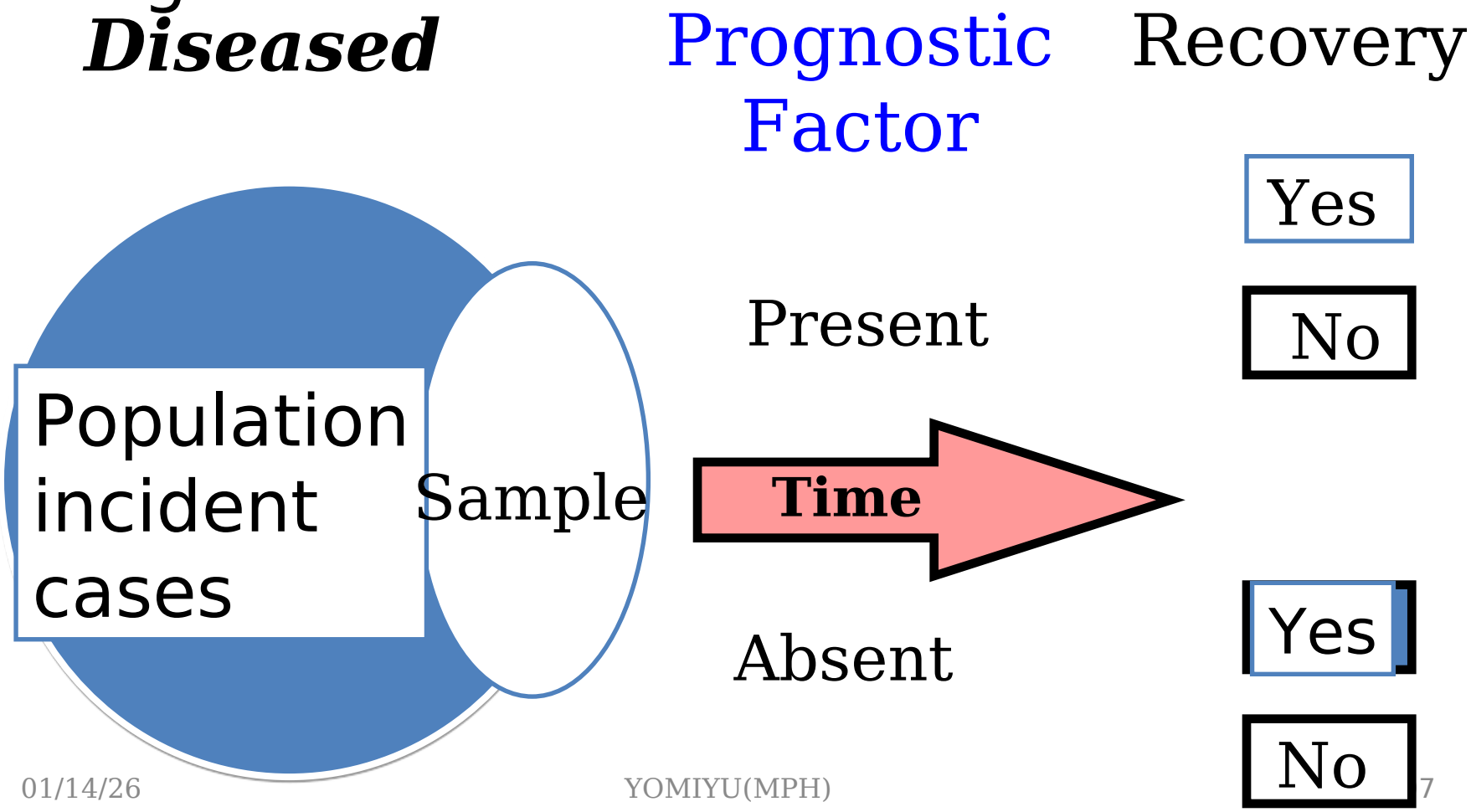


Prognostic Cohort Study

- Follow diseased cohort to assess factors associated with outcome (recovery or death)
- Goal is to identify explanatory/prognostic factors of those factors helped to the dev't of the disease.

Prognostic Cohort

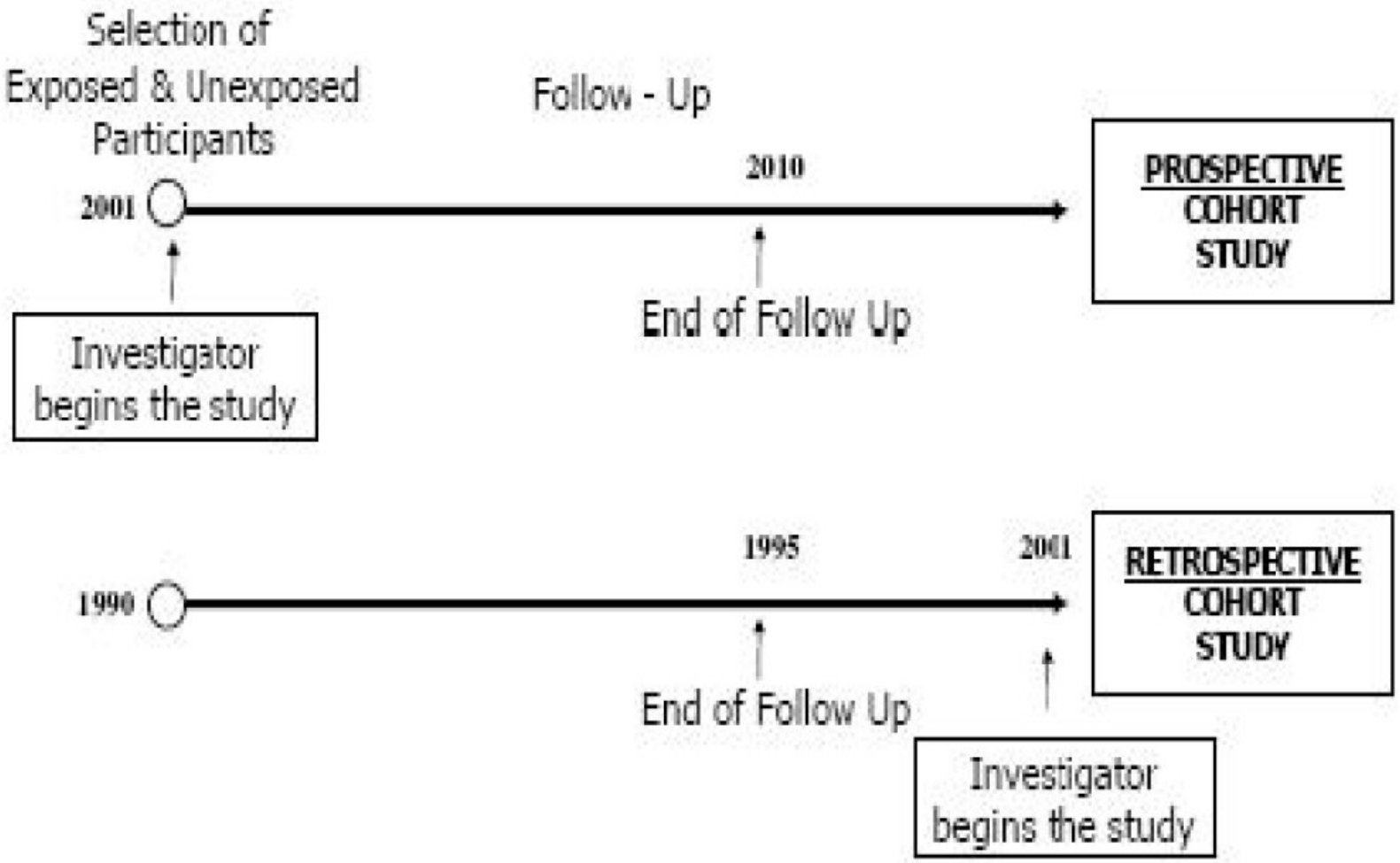
- Fig.



Types (3)

- Prospective vs. Retrospective (Concurrent vs. Non-concurrent)
  - Depending on temporal relationship between initiation of the study /timing of data collection/ and occurrence of the disease
  - Cohort studies have been called prospective studies, but this terminology is confusing and should be avoided
    - the term “prospective” refers to the timing of data collection and not to the relationship between exposure and effect
  - Thus there can be both prospective and retrospective cohort studies

Retrospective Vs Prospective C ohort Study



Prospective cohort studies

- Exposure and outcome data is collected after start of the study
  - cohorts Identified in the present
  - exposure status or possible explanatory/prognostic factors determined in the present
  - Cohorts followed-up to identify outcome
  - Ascertainment of outcome done in future

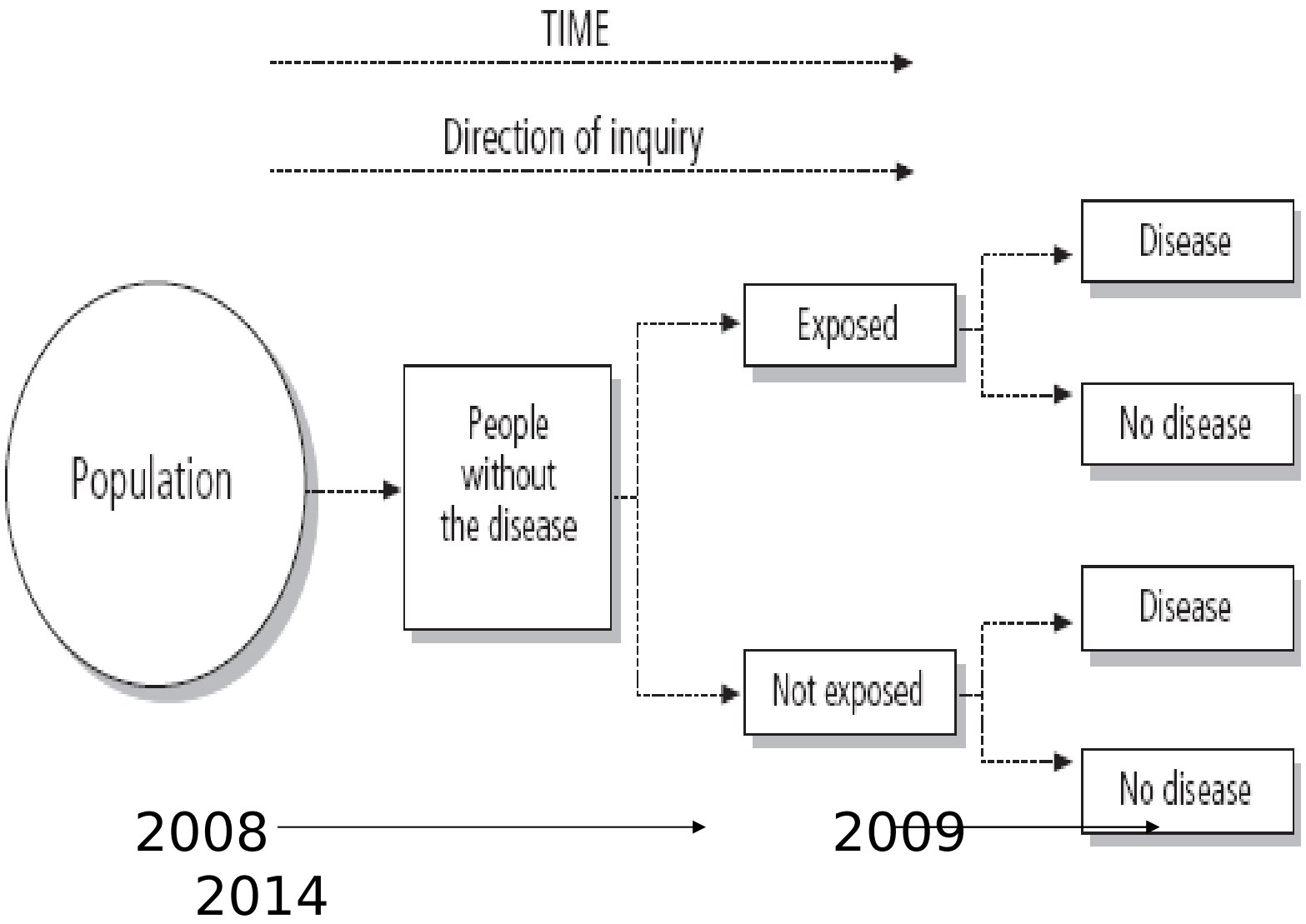


Fig. 2. Design of prospective cohort studies

Retrospective cohort (1)

- all the exposure and effect data have been collected before the actual study begins
- This type of investigation is called a historical cohort study
- Conduct
  - Identify cohort in the past using records/databases
  - Determine exposure or prognostic factors in the past using again records or databases then
  - Identify outcome in past or present or future (in case of mixed cohort)

Retrospective cohort (2)

- Costs can occasionally be reduced by using a historical cohort (identified on the basis of records of previous exposure)
- This sort of design is relatively common for studies of cancer related to occupational exposures
- For example, records of military personnel exposure to radioactive fall-out at nuclear bomb testing sites have been used to examine the possible causal role of fall-out in the development of cancer over the past 30 years

Retrospective vs. Prospective

- Can be conducted more quickly and cheaply
  - All relevant events have occurred
  - Efficient for disease with long latency periods
- Depend on availability of routine data
  - Incompleteness
  - Lack of data on confounding variables

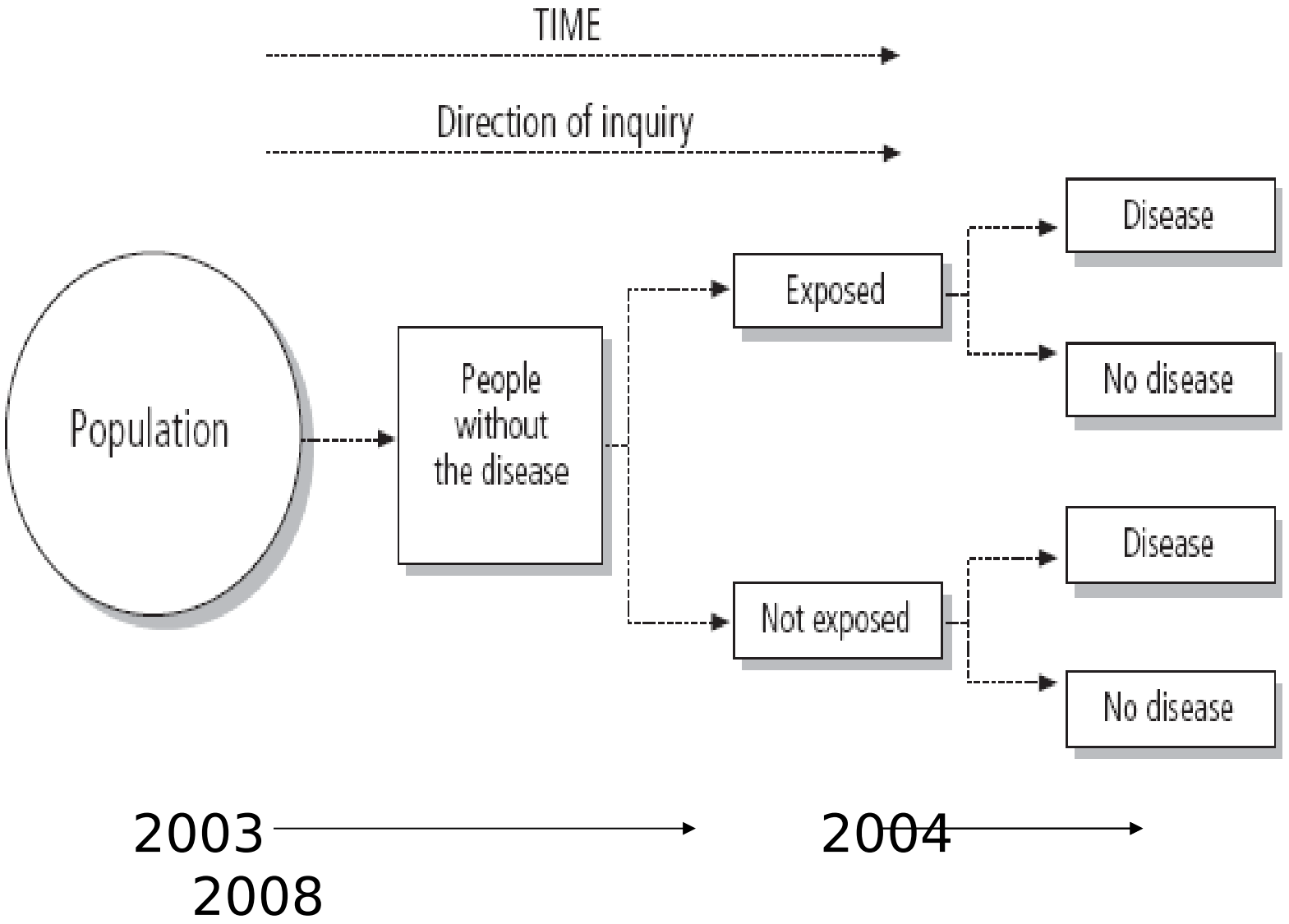


Fig. 3. Design of retrospective cohort studies

Time and Cohort Studies



Cohort Assembled — Follow-up = Historical (retrospective)

Prospective = Cohort Assembled — Follow-up

Mixed Cohort

Design and data collection (2)

  - Outcome must not be rare in exposed
  - Attributable risk must be high
  - Accessible and compliant subjects
- Selection of Non-exposed Group
  - Similar to exposed
  - Control for confounding factors

Design and data collection (1)

1. Define and identify cohorts

1.1. Identify population at risk

- Selection of Exposed Population
  - Depends on research question
  - Depends on frequency of exposure
    - Common exposures: general population
    - Rare exposures: selected groups

Design and data collection (3)

1.2. Screen identified subjects for the disease and omit the prevalent cases

2. Define, assess, identify and classify exposure

3. Follow-up and ascertain outcome

   - Timing of outcome events-case definition

Selection of Non-Exposed (controls)

- Need to be comparable to the c haracteristics of the exposed population.

Exposed groups Characteristics	Controls Characteristics
• Smokers	• Not-Smokers
• Largely Educated	• Largely Educated
•Urban residents	•Urban residents
•Male	•Male

Exposure Ascertainment

Pre-existing records: from hospital, employers record..

- can make available information for large number of individuals!
- relatively inexpensive
- allow objective and unbiased determination of exposure status

But...

- information on exposure level may be insufficient
- may not contain adequate information on potential confounders

Survey:

- enables to record exposure information that are not routinely recorded, particularly lifestyle factors

But...

- potential for information bias, particularly recall

Outcome Ascertainment

- Obtain complete, comparable and unbiased information on the subsequent health experience of every study subject via:
  - routine surveillance
  - periodic health examination
  - hospital records
  - death certificate. etc.
- Using a firm outcome criteri a and standard diagnostic procedure that are equally app lied for exposed and non-exposed individuals

Sampling

Sample size - for test of significant difference between two proportions, the following formula can be used:

$$n = \frac{(Z_{\alpha} + Z_{2\beta})^2 \{ p_1(1- p_1) + p_2(1- p_2) \}}{(p_1 - p_2)^2}$$

Parameters:

n - size of sample in each group

P_1, P_2–estimated population prevalence in the comparison groups

β = 1- Power (the probability that if the two proportions differ the test will produce a significant difference)

– Usually a power of 80% is used

Analysis

RR=INCIDENCE EXP/INC NONEXP

- Presence of association
  - Population
    - RR=1 – no association; RR<1 – negative association; RR>1 – positive association
    - AR, PAR & %
  - Sample
    - P-value≤0.05 – statistically significant association
      - RR≠1
    - Statistical methods – survival analysis
- Strength of association
  - Weak – RR close to zero; Strong – RR far from zero

Measures in Cohort Studies

	Develops Disease	Doesn't Develop Disease	Totals	Incidence Rates of Disease
Exposed	a	b	a + b	$\frac{a}{a+b}$
Not Exposed	c	d	c + d	$\frac{c}{c+d}$

Relative Risk (RR) = $I_{exp} / I_{non\text{-}exp} = [a/(a+b)] / [c/(c+d)]$

Advantages

- elucidates temporal relationship between exposure and outcome
- allows direct measurement of the risk of developing disease
- it is possible to examine a range of outcomes
- Of particular value when exposure is rare
- Minimize selection bias - prospective

Disadvantages (1)

- Time-consuming - prospective
  - require long periods of follow-up since disease may occur a long time after exposure
  - For example, the induction period for leukaemia or thyroid cancer caused by radiation is many years and it is necessary to follow up study participants for a long time
  - Many exposures investigated are long-term in nature and accurate information about them requires data collection over long periods

Disadvantages (2)

- Expensive - prospective
  - Often involve following large number of individuals for many years
  - Exposure assessment
  - Therefore, conducted after hypothesized relationship has been evaluated in case-control design
- Incompleteness of records – retrospective
- Loss-to-follow-up – bias
- Inefficient for rare disease

Bias and confounding

- Selection bias – less
- Misclassification of exposure &/or outcome
  - Due to problems in accuracy and completeness of information
  - Differential and non-differential
- Loss to follow-up – major source of bias
  - <30-40%
  - Solution: minimize, obtain data about losses to follow-up, extreme analysis
- Non-participation
- Confounding-less

Selection of study design

- Depends on
  - Hypothesis being tested
  - Resources available
  - Current state of knowledge

Intervention studies

Learning objectives

- Define common terminologies of interventions studies
- Describe the concept of different types of clinical trial
- Understand the concept of community trial
- Be able to conduct clinical trial

Discussion

- What are the unique features of clinical trial?
- What is randomization and its purpose?
- Elaborate the concept of community trial?

INTERVENTION STUDIES

- Investigator determines who is exposed, ideally using random methods
- Investigator allocates the exposure and follows for an outcome
- Types of interventional studies include
  - Randomized Clinical Trials
  - Field Trials
  - Community Intervention Trials
  - Quasi-experimental Studies

What is an experimental study?

- Randomized controlled trials are sub-types of cohort studies in which exposure (i.e., treatment) is randomly assigned by the investigator (or by some other, observable phenomenon)
- Have a long history in clinical medicine
- Although experimental studies come in many types, principles are the same and clinical trials dominate the field

When to choose an experimental design?

- Experimental studies are generally reserved for relatively "mature" research questions.

A lot of ground work has to be done before embarking on a clinical trial.

What is a clinical trial?

- A clinical trial is a prospective study evaluating the effect and value of intervention(s) in human beings under pre-specified conditions.
- A controlled clinical trial is a prospective study comparing the effect and value of intervention(s) against a control in human beings.

- The clinical trial is the most definitive tool for evaluation of the applicability of clinical research.
- It represents a key research activity with the potential to improve the quality of health care and control costs through careful comparison of alternative treatments.

When might a RCT be indicated?

- Exposure is a modifiable factor which persons will *let you modify,* e.g., taking a pill, trying a different diet
- A particular exposure may have an influence on multiple outcomes of tremendous importance
- In experimental trials, in contrast with other epidemiologic study designs we have discussed, we are *doing something* to participants so we have to be certain that, *first, we do no harm*

Clinical Trial Phases

- Phase I: clinical pharmacology and toxicity
- Phase II: Initial Assessment of Efficacy
- Phase III: Full-scale Evaluation of Treatment Efficacy
- Phase IV: Postmarking Surveillance

Ref: Nottage M and Siu LL. Principles of clinical trial design. Journal of Clinical Oncology 20:42s-46s, 2002.

Classification of Intervention Studies:
Based on Trial Objective

- **Phase I -** trial on small subjects to test a new drug with small dosage to determine the toxic effect.
  - Scale: 20-80 healthy individuals
- **Phase II -** trial on small group to determine the therapeutic effect.
  - Scale: 100-200 patients
- **Phase III -** study on large population to test effectiveness.
  - ***Randomised Controlled Trial (RCT)***
  - Often it is multi-centered …… licensing
- **Phase IV**

– Post marketing surveillance

– Long term prospective assessment of effects & side-effects

Conducting Trials

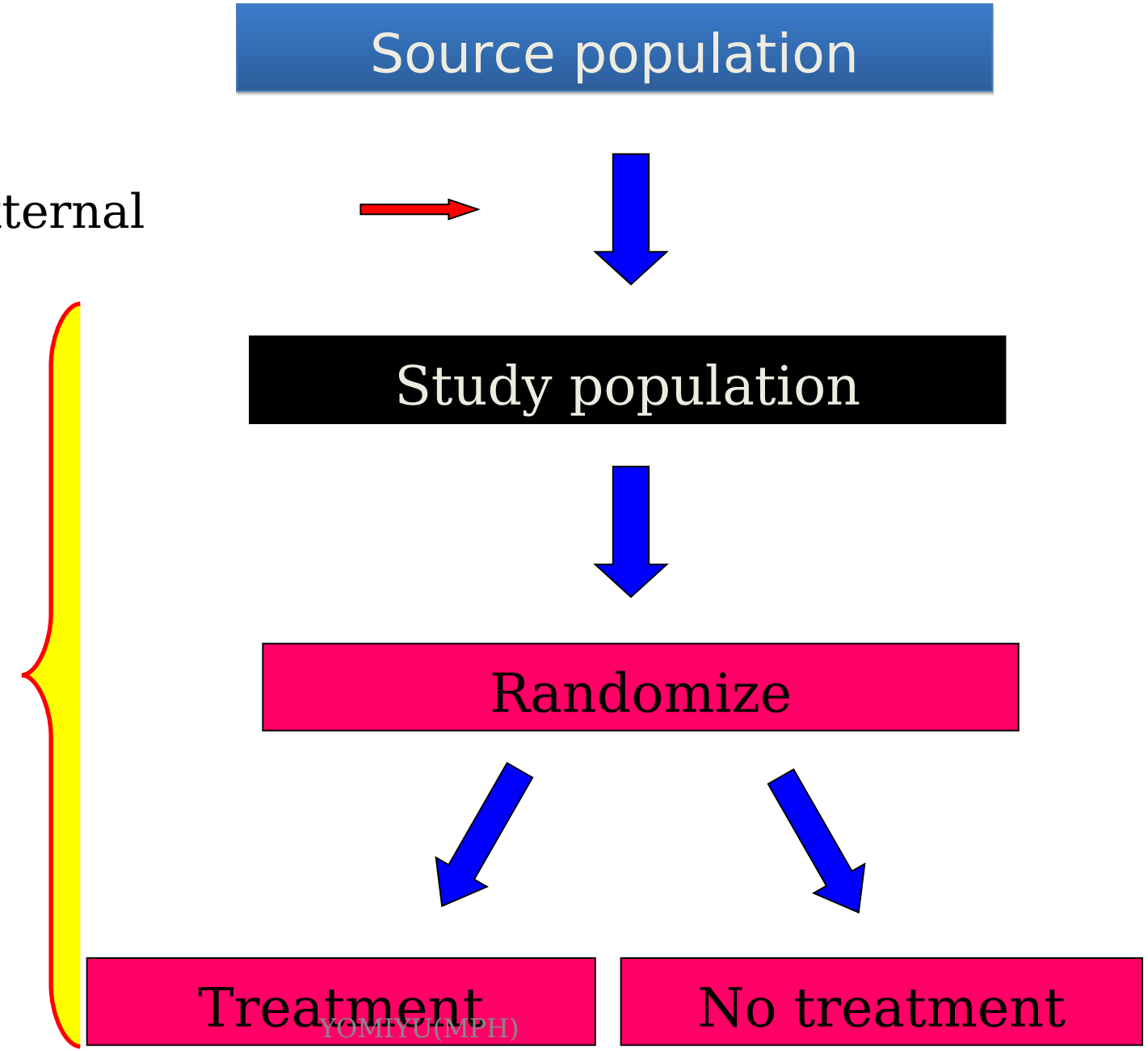
1. Selecting participants
2. Measure baseline characteristics and describe sample
3. Randomizing
4. Apply intervention
5. Follow-up and adherence to protocol
6. Measuring outcome
7. Analysis

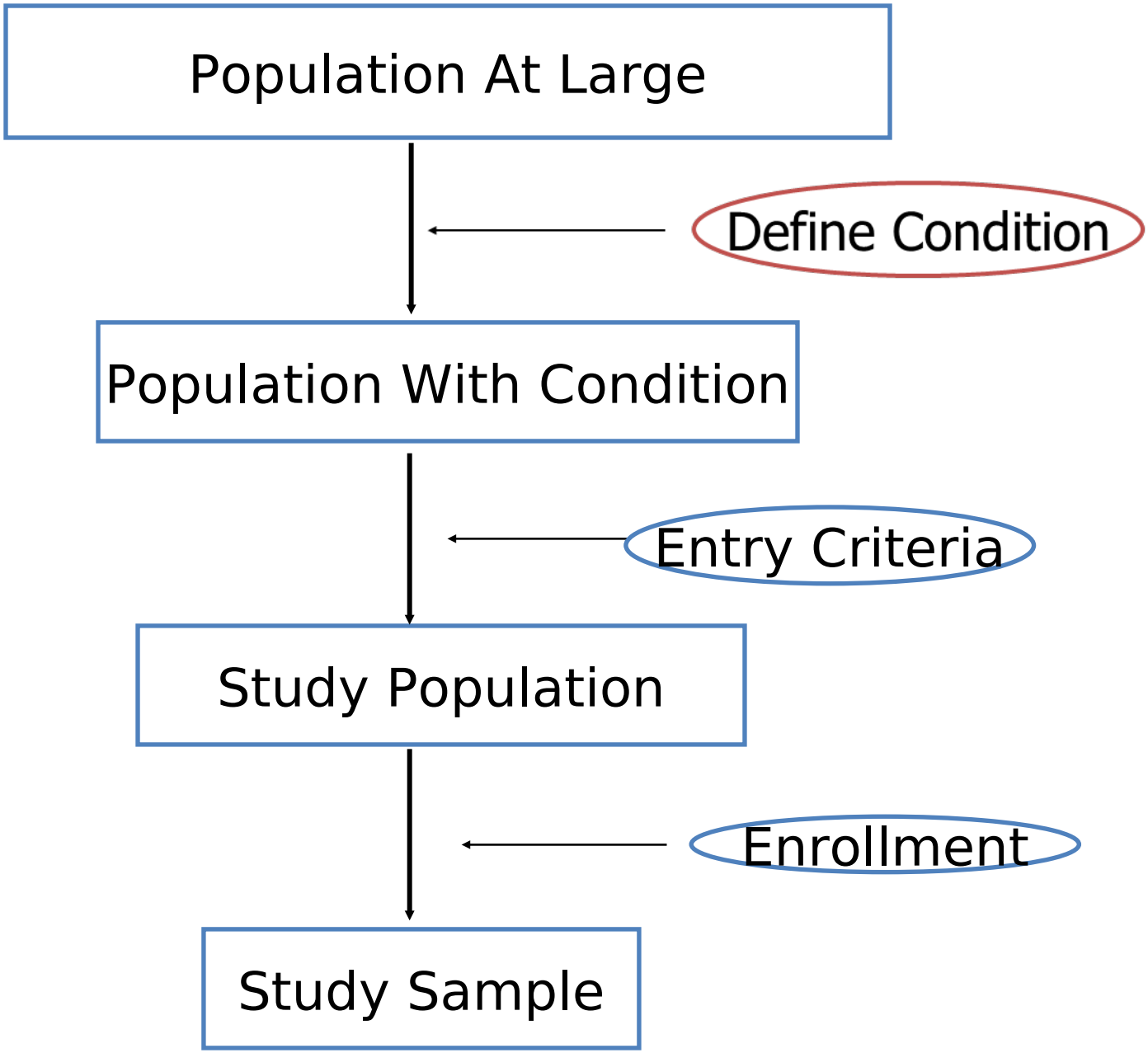
1. Selection of Participants

- Terminology
  - Target population
    - People to which findings will be generalized
    - Population from which sample is selected
  - Study Sample
    - Subset of population available to study
    - Subjects in the study
  - Selecting subjects
    - Establish inclusion/exclusion criteria
    - Sample size

threats to external validity

threats to internal validity




Population At Large
Define Condition
Population With Condition
Entry Criteria
Study Population
Enrollment
Study Sample

Participants

- Inclusion criteria
  - Define main characteristics of target population that are relevant to research question
    - Demographic characteristics
      - E.g., adults, aged 20-69
    - Clinical characteristics
      - E.g., in good health
    - Geographic characteristics
      - E.g., living in northern Ethiopia
    - Temporal characteristics
      - E.g., inception period Jan 1, 2003 to Dec 31, 2003

Participants

- Exclusion criteria
  - Subsets of people meeting inclusion criteria (potentially suitable for research question) except for characteristics that might interfere with quality of data etc.
  - High likelihood of being lost to follow-up
    - E.g., transients
  - Inability to provide good data
    - E.g., language barrier or cognitive incapacity
  - High risk for side effects
    - E.g., pregnant, lactating
  - Unethical to withhold study treatment
    - E.g., severe depression

Exclusion Criteria

- Five main reasons for exclusion from clinical trial
  - Safety concerns (susceptibility to adverse effects of active treatment)
  - Unethical to withhold treatment (tx so beneficial for some → not acceptable to assign placebo
  - Active treatment unlikely to be effective
  - Unlikely to adhere to treatment
  - Unlikely to provide outcome information (e.g., die or move before study completion)

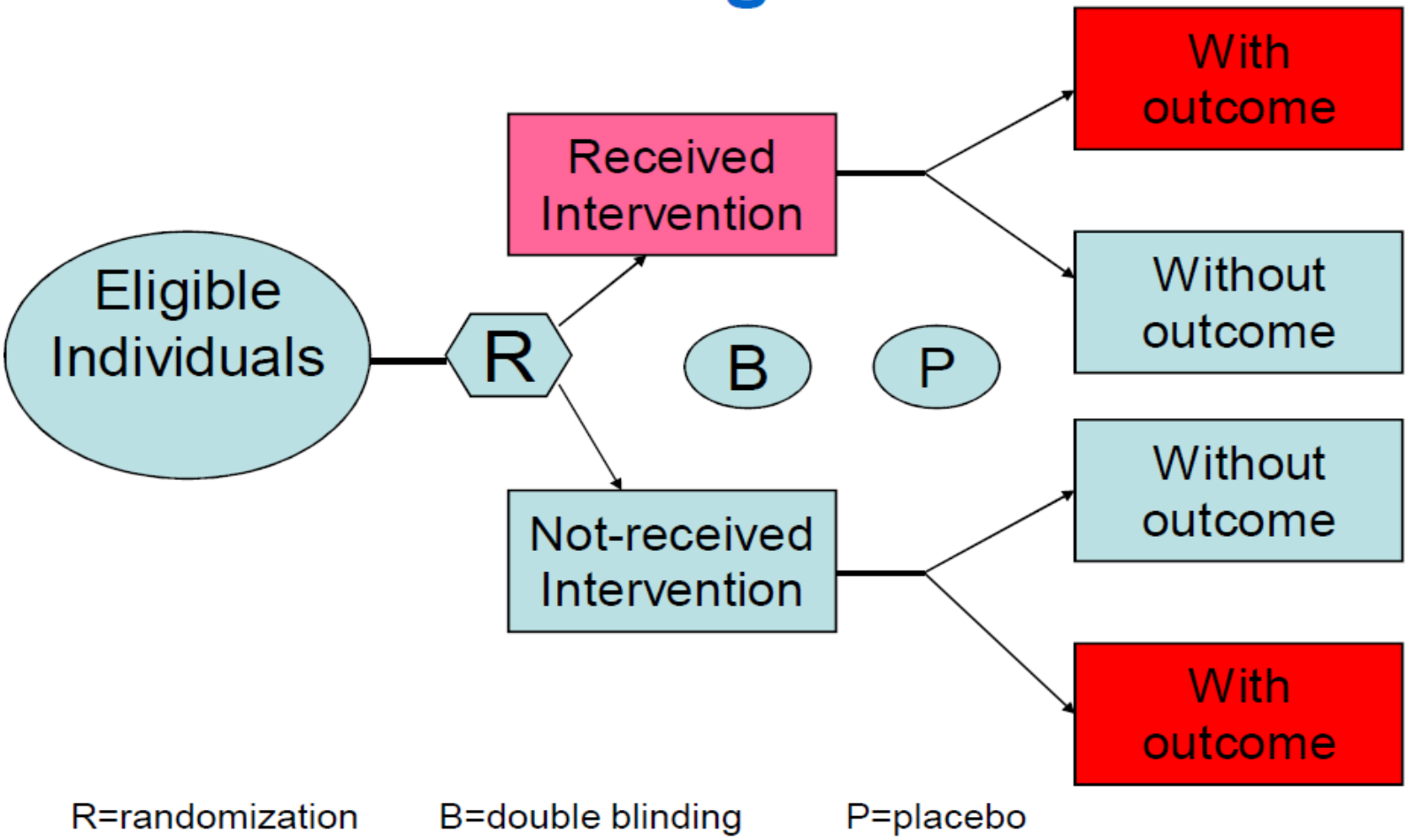
2. Baseline Characteristics

- Need enough information to track subjects
  - Contact persons; address, etc.)
- Description of participants
  - Aid in assessing generalizability (e.g., gender, age, disease severity, etc.)
- Risk factors for outcome or to define subgroups
  - E.g., smoking status, smoking status of spouse
- Measure of “outcome” variable
  - E.g., if pain is “outcome”, need baseline pain

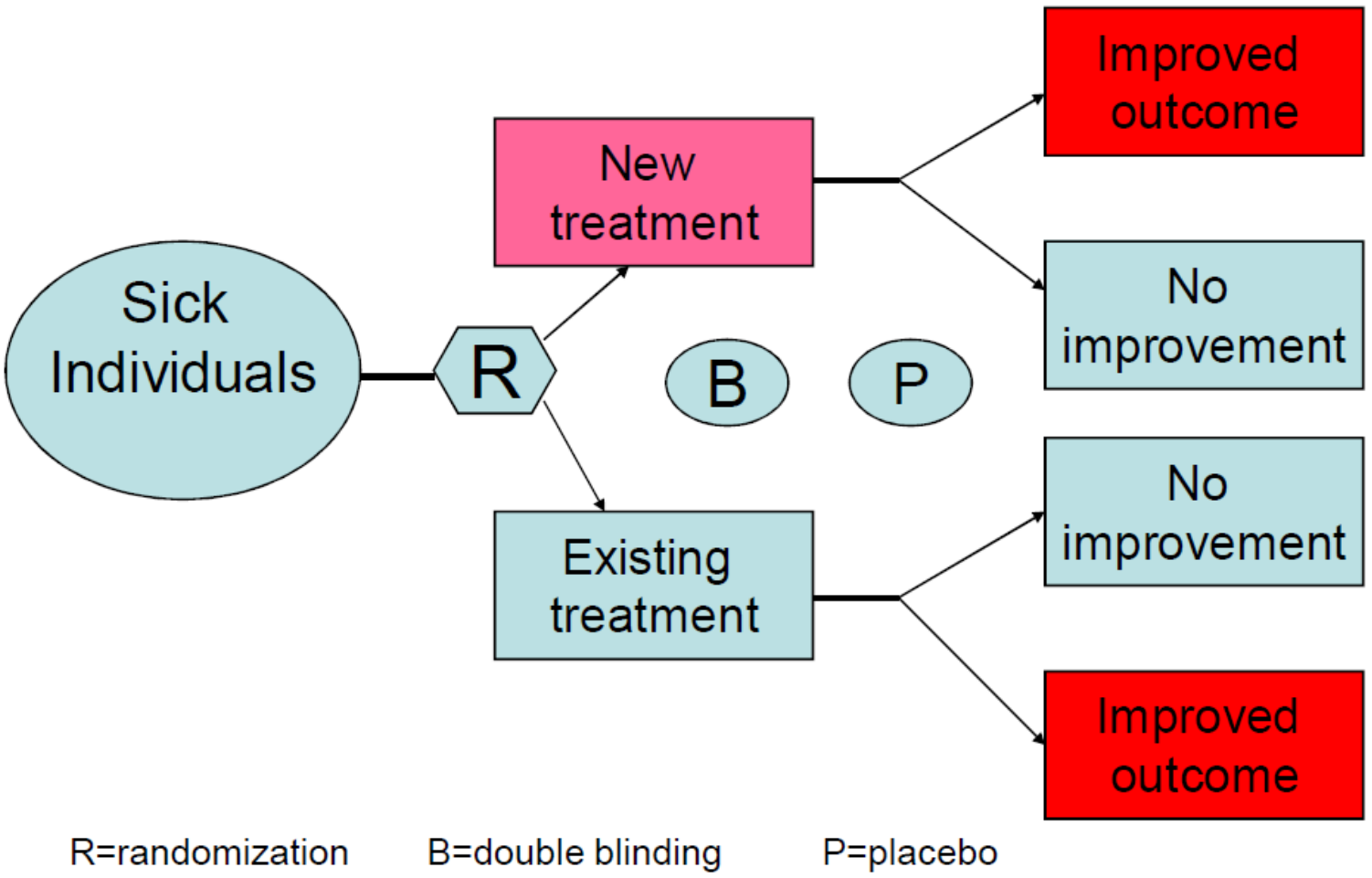
3. Randomization

- Should be done at the last possible moment, *after* eligibility criteria has been determined and informed consent has been obtained
- Important to conceal randomization scheme from attending health care providers

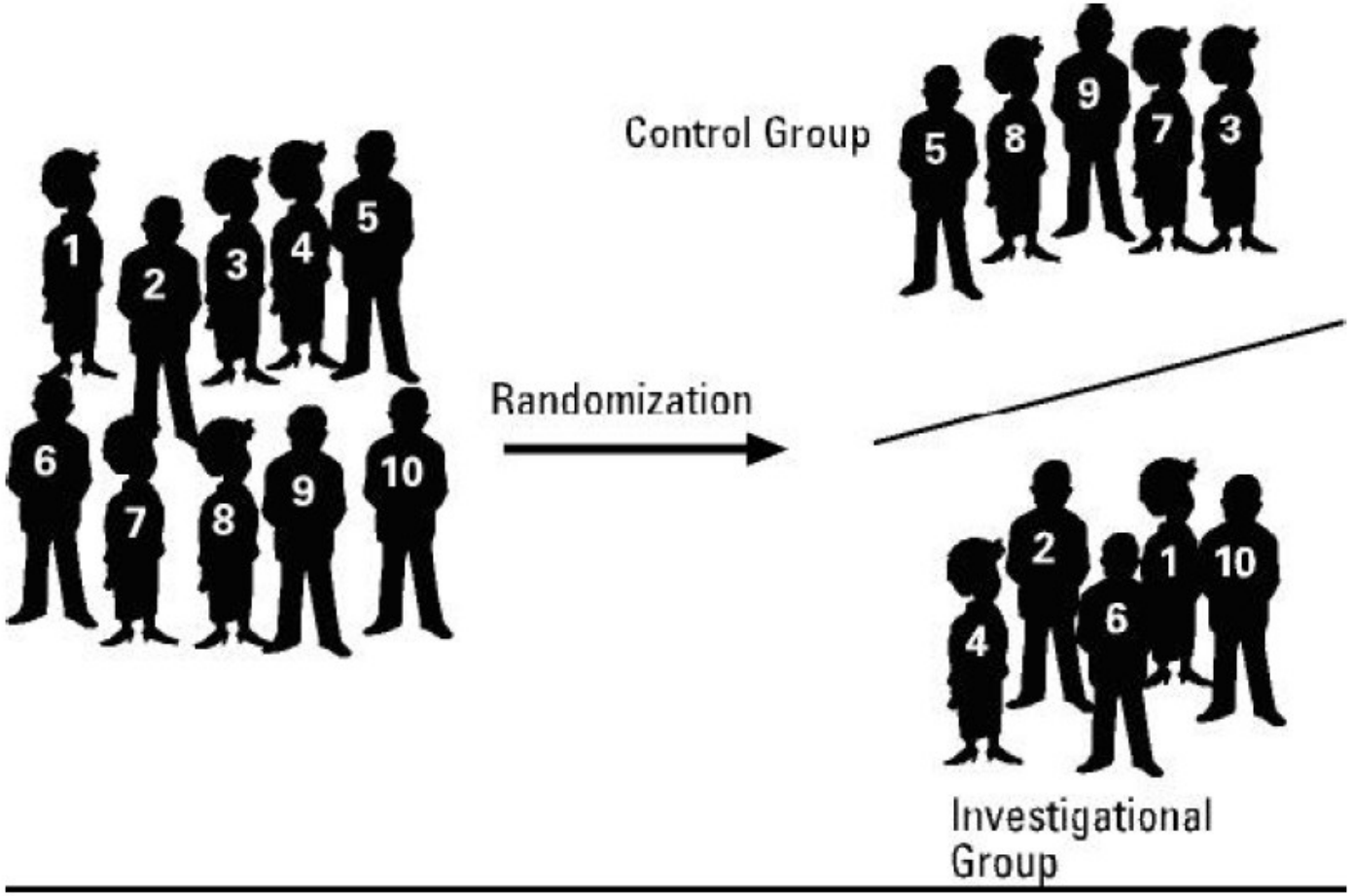
Intervention (Experimental) Design



Therapeutic trial



Random Allocation



Advantages of Randomization

- Removes the potential of bias in the allocation of participants.
- Prevents confounding
- produce comparison groups

Methods of randomization

- Fixed allocation Randomization
  - simple Randomization
  - Stratified Randomization

Constrained randomization

Stratified Randomization

- Select factor(s) of interest
- Stratify (divide) group by that factor
- Randomize the appropriate proportion of each group into your treatment groups
- Increases the likelihood that your treatment groups will be comparable on that factor

Stratified Randomization

- E.g., age is important in treatment response
  - 100 in sample → want 50 per group
  - Say 20% of sample are >60 (high risk group)
  - Randomly select 10 people over 60 for each group
  - Then randomly select 40 people under 60 for each group

4. Applying Intervention

- Intervention strategy compared with
  - Placebo
  - Standard treatment
- May have more than one comparison group

1. Placebo

a. A biologically inert intervention which is used to elicit any non-specific psychological effects of the test intervention.

b. Designed to resemble the test intervention as closely as possible, except for the presumably active component.

c. May not be justified if an intervention of known effectiveness is already in general use.

2. Using a specific alternative intervention

a. For example, comparing Vaccine A versus Vaccine B.

b. Sometimes used to test a hypothesis of equal efficacy (or bioequivalence), possibly comparing a new intervention to the best existing alternative.

Blinding

- Randomization
  - Control for confounding bias at baseline
  - Does not control for confounding during follow-up
    - E.g., differential attention to subjects in treatment arm (a type of co-intervention)
  - Does not control for information bias
- Blinding
  - Controls for
    - Information bias (e.g., observer bias)
    - Reduce loss to follow up (reduce selection bias)

Blind studies

Single — The patients do not know which treatment they receive

Double blind — The patient and the observer or the physician do not know

Triple blind — The patient, the observer and the analyst do not know

Why use blinding?

- Use blinding to eliminate the effect of knowledge of treatment - which introduces bias
- If a patient is aware of assignment, knowledge of therapy may affect outcome (use a *placebo)*
- Observer may differentially recognize and record information regarding the patient based on the treatment group of the subject

Intervention

- All patients must receive ***identical*** care part from the intervention …
  - Tablets must look/taste the same
  - Equal time must be given to all patients
  - No hidden intervention, e.g. doctors or nurse
  - Intervention should last for appropriate time
  - Control Intervention: Standard care *vs* Placebo

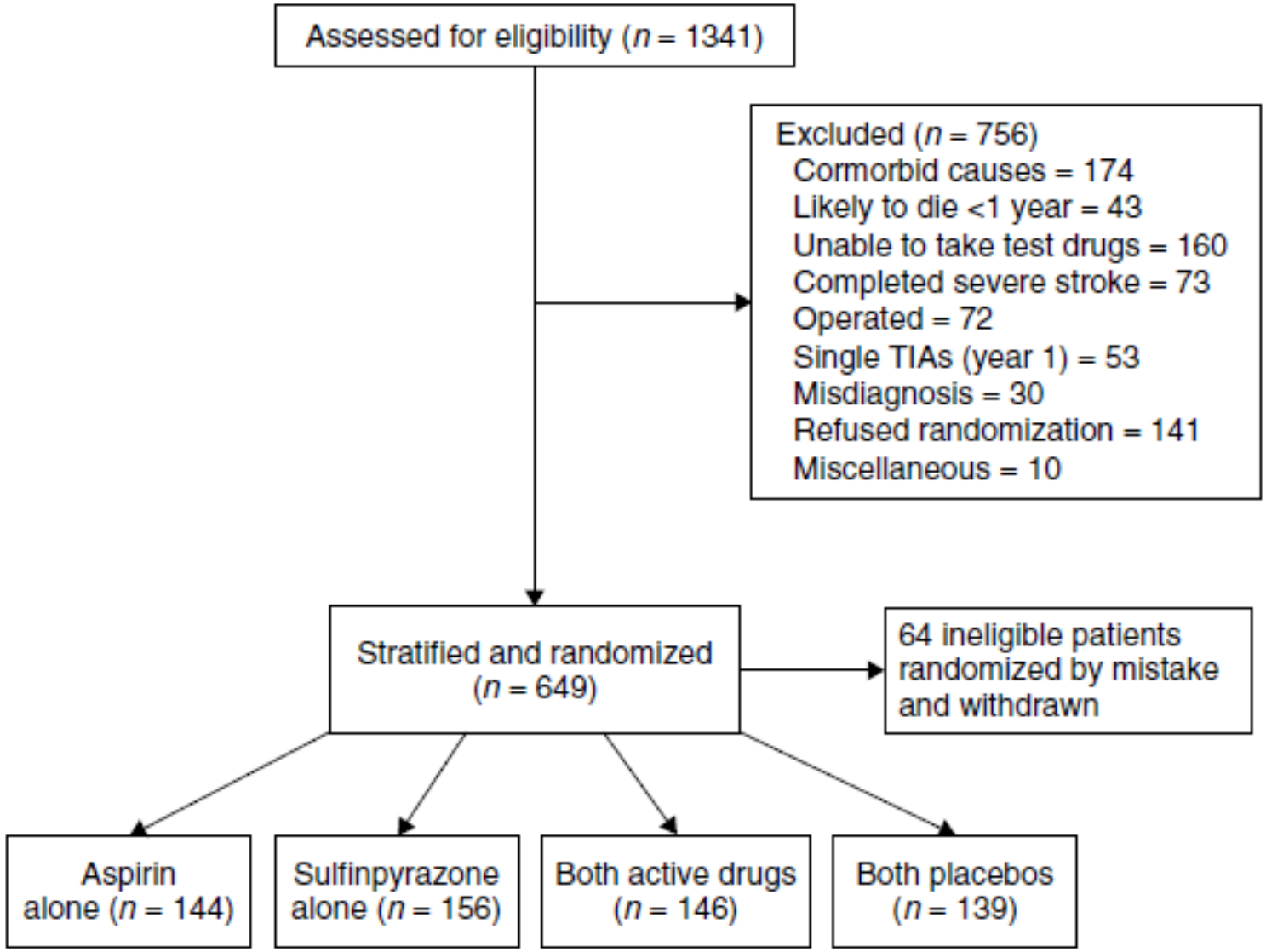


FIGURE 5–3 Recent Recurrent Presumed Cerebral Emboli (RRPCE) patient-flow diagram after allocation.

Ethical Considerations

- Risks *vs* benefits
- Comparison: Standard care *vs* placebo
- Ethical approval
- Informed consent & confidentiality
- Freedom to withdraw
- Duty of care
- Stopping/Monitoring
- Reporting findings
- Quality: 'Poor' quality research is unethical!

Practical Considerations

- Supplies of drugs/placebos
- Compliance of clinicians/nurses & patients:
  - Keep data required to a minimum
  - Keep forms simple & logical
  - Keep people informed of progress, e.g. newsletters,
- Patient Information Leaflet
  - Clear & easily understood
  - Voluntary nature, free to withdraw
  - Aim of RCT
  - Interventions, incl. Side effects
- Consent form

Monitoring Trials

- An independent group must be established.
- Purpose:
  - Safety of subjects
  - Maintaining the quality of the study
  - Maintain objectivity

5. Follow-up and Adherence

- Ideally
  - All subjects adhere to treatment regimens; do not seek additional treatment; do not drop out, die, move or have to be withdrawn from study; attend follow-up sessions and provide outcome data

6. Measuring Outcome

- Outcome should be
  - Clinically relevant
  - Feasible
    - E.g., choose outcome that is sufficiently common for time and number of subjects
    - Able to be measured accurately and precisely
    - Not too costly
  - Above all, valid, reliable and sensitive to change

Timing of the outcome events/endpoints

- Rigorously defined endpoints is among the characteristics of experimental studies
- Endpoints are the quantitative or qualitative measurements implied by the objectives
- The best endpoints to use depend on specific interventional objectives

Surrogate endpoints

- Sometimes called intermediate, surrogate marker or replacement endpoint
- Is one that is measured in place of biologically definitive or clinically most meaningful endpoints
- Tracks the progress or extent of the disease
- Needs to be associated with a definite clinical outcome so that it can be a reliable measurement
- Investigators choose a surrogate when the definitive endpoint is inaccessible due to cost, time, or difficulty of measurement

Example of Surrogate endpoints

Disease	Definite endpoints	Surrogate endpoints
HIV infection	AIDS (or Death)	CD_4^+ count
Cancer	Mortality	Tumor size reduction
Colon cancer	Disease progression	CEA Level
Prostate cancer	Disease progression	PSA Level
Cardiovascular disease	Hemorrhagic stroke Myocardial infarction	Blood pressure Cholesterol level
Glaucoma	Vision loss	Intraocular pressure

Analysis

- Analyze according to which treatment the patient was randomized to (which treatment was intended?)
  - “intention to treat” analysis
- Or according to which treatment they actually received?
  - “per protocol” analysis

ITT---cont

- Key points
- ❖Use every subject who was randomized according to randomized treatment assignment
- ❖Ignore noncompliance, protocol deviations, withdrawal, and anything that happens after randomization
- ***As randomized, so analyzed***

Prepare a 2X2 table

	Cases/with	Non-cases
Exposed	a	b
Non-exposed	c	d
All	a + c	b + d

N = a+b+c+d

Case-control study:

OR = ad / bc

Cohort /Exp`tal study:

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Scenario

- In a randomized clinical trial participants with untreated mild hypertension were randomly assigned to either standard anti-hypertensive treatment or to a new test treatment regimen
- Subjects were followed for heart attack during a two-year follow-up period
- Among the 950 subjects in the group with standard treatment 190 developed a heart attack during follow-up
- Among the 948 subjects in the test treatment group 96 developed a heart attack during follow-up

Experimental designs are chosen when:-

- The research question cannot be answered by observational studies
- Earlier observational studies have not answered the research question
- Existing knowledge is not sufficient to determine clinical or public health policy
- An experiment is likely to provide an important extension of this knowledge

Experimental studies: Advantages

- The major advantage of experimental studies lie in the strength of causal inference that can be made.
  - it is very difficult to make causal inferences based on observational studies.
- Experimental studies offer the best design for controlling confounding variables.
- Randomised Controlled Trials (RCTs)
  - Gold standard for epidemiologic research

Community Trials

- Epidemiological experiment where the units of analysis are entire communities or groups of people
  - E.g., Cities, towns, villages, schools, occupational settings
  - Have 1 of 2 major objectives:
    - (a) Test etiologic hypotheses about health-related exposures and outcomes
    - (b) Evaluate the impact of health-related programs, projects, or campaigns on communities (EVALUATION STUDIES)

Overall study protocol

- Develop a research hypothesis
- Select communities to be studied
- Collect baseline data on study outcomes
- Allocate communities into experimental and control groups
- Apply intervention
- Assess outcomes during follow-up period

Examples of community-level interventions

- Screening or immunization programs delivered to residents of a geographic area
- Health promotion programs delivered to towns, schools
- Services provided to primary care practice populations

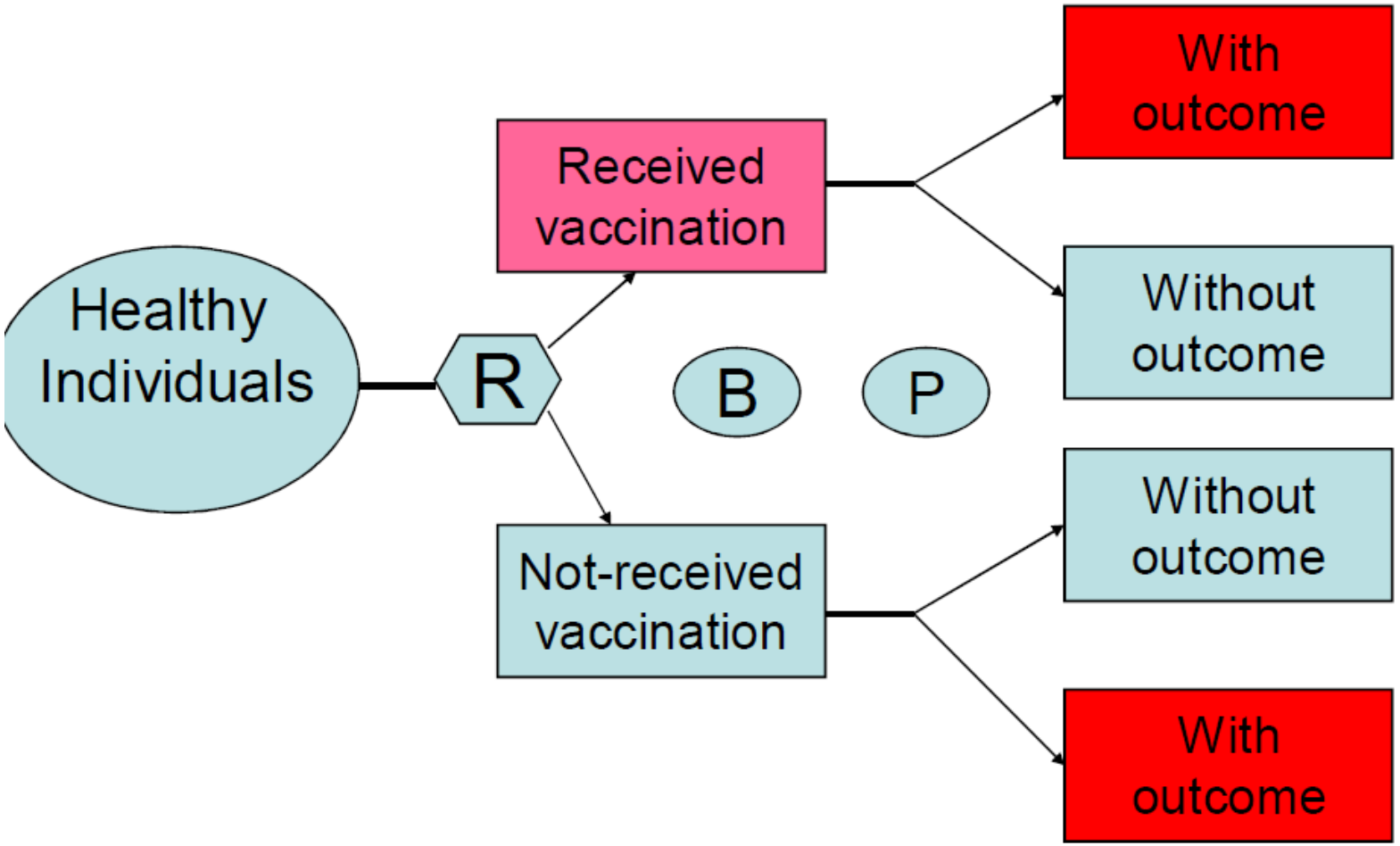
Examples of group interventions

- Educational interventions
- Group psychological interventions
- Nutritional interventions
- Environmental sanitation interventions:
  - delivered to household, village etc
  - latrines, dietary supplements

Reasons for carrying out evaluations at group or cluster level

- More appropriate for interventions delivered to groups
- Individual randomization may not be feasible because all members of group are treated same way
- Individual randomization, although feasible, may result in substantial “contamination”

Prevention trial



R=randomization B=double blinding P=placebo

Community based trials

All schools in an area

20 reasonably isolated schools

10 schools randomised to receive sex education

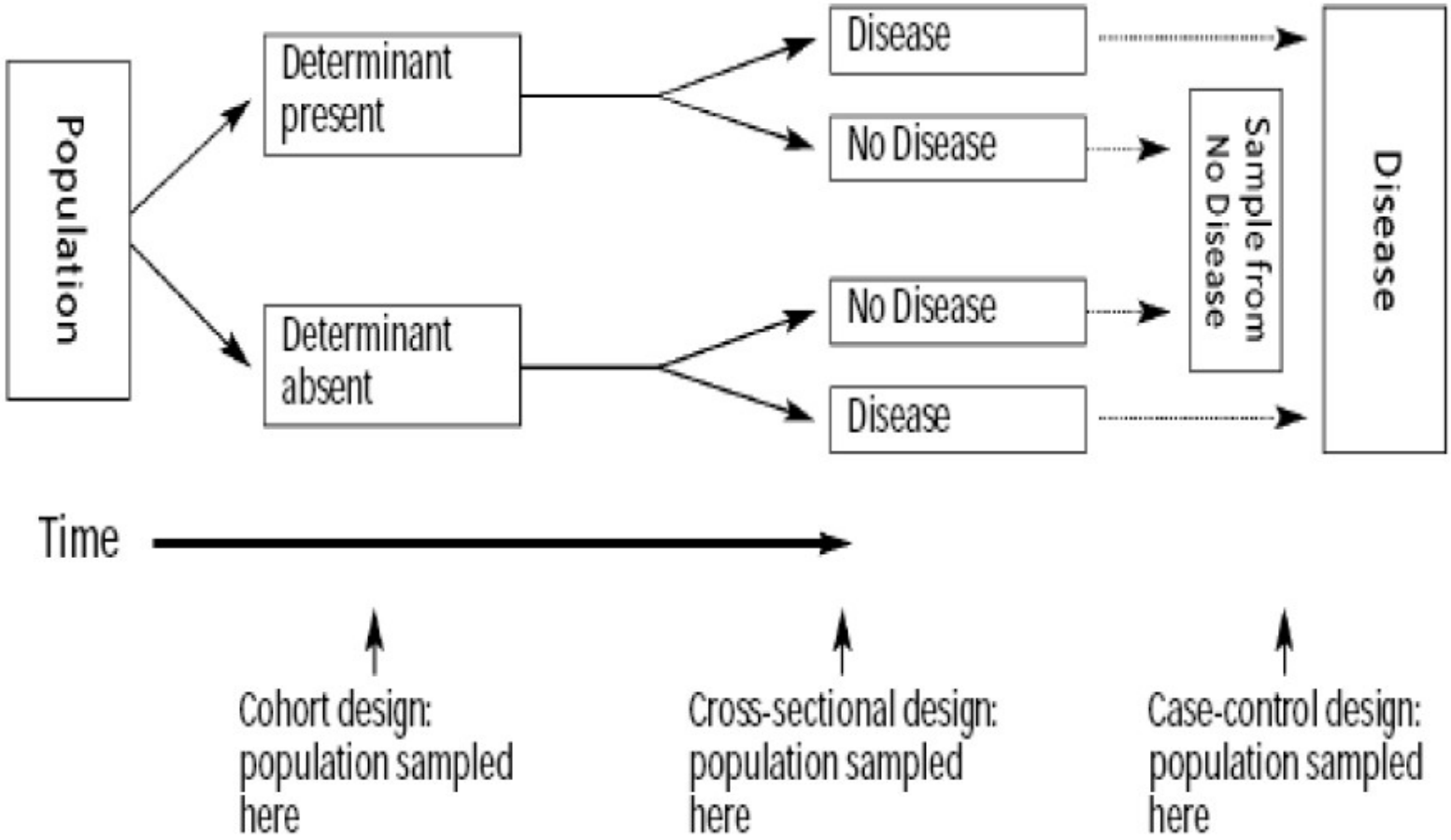
How many new HIV infections at age 18

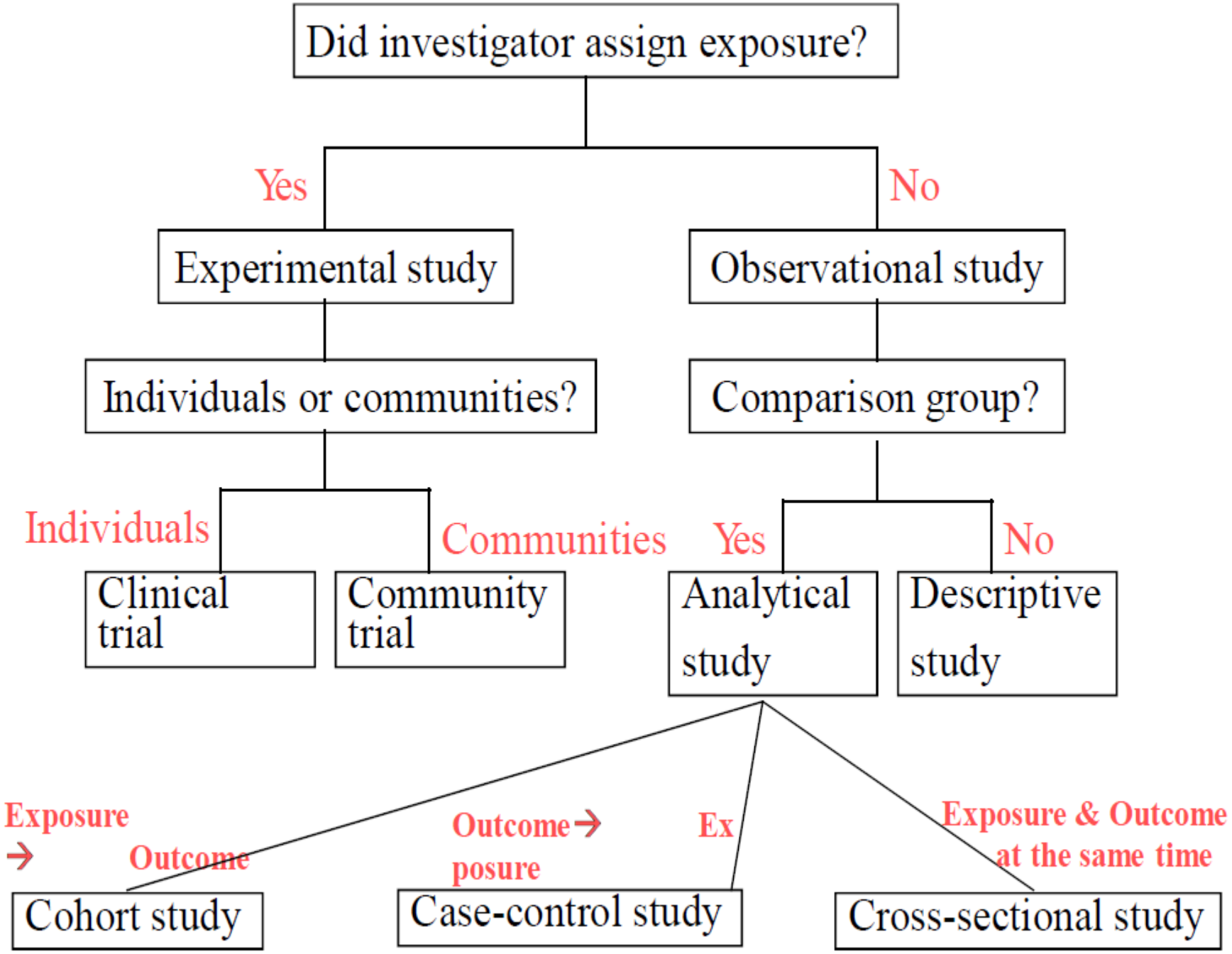
10 schools do not receive sex education immediately

How many new HIV infections at age 18

In summary

Overview of Analytical Studies




Did investigator assign exposure?
Yes
No
Experimental study
Observational study
Individuals or communities?
Comparison group?
Individuals
Communities
Yes
No
Clinical trial
Community trial
Analytical study
Descriptive study
Exposure → Outcome
Outcome → posure
Ex
Exposure & Outcome at the same time
Cohort study
Case-control study
Cross-sectional study

Strength of Evidence!

