Epidemiology

- Introduction
- Types of Studies

Definition of Epidemiology

"The STUDY of the DISTRIBUTION and DETERMINANTS of HEALTH-RELATED STATES in specified POPULATIONS, and the application of this study to the CONTROL of health problems."

Key words

- Study
- Distribution
- Determinants
- Health-related states
- Populations
- Control

Epidemiology key words: Study

- Epidemiology is the basic science of public health.
- It's a highly quantitative discipline based on principles of statistics and research methodologies.

Epidemiology key words: **Distribution**

- Epidemiologists study the distribution of frequencies and patterns of health events within groups in a population.
- They use descriptive epidemiology, which characterizes health events in terms of time, place, and person.
  - Questions: who, what, where, when

Epidemiology key words: **Determinants**

- Epidemiologists also attempt to search for causes or factors that are associated with increased risk or probability of disease.

- This type of epidemiology is referred to as analytical epidemiology

  - Questions: how and why

Epidemiologic Process

***WHO* was affected?**

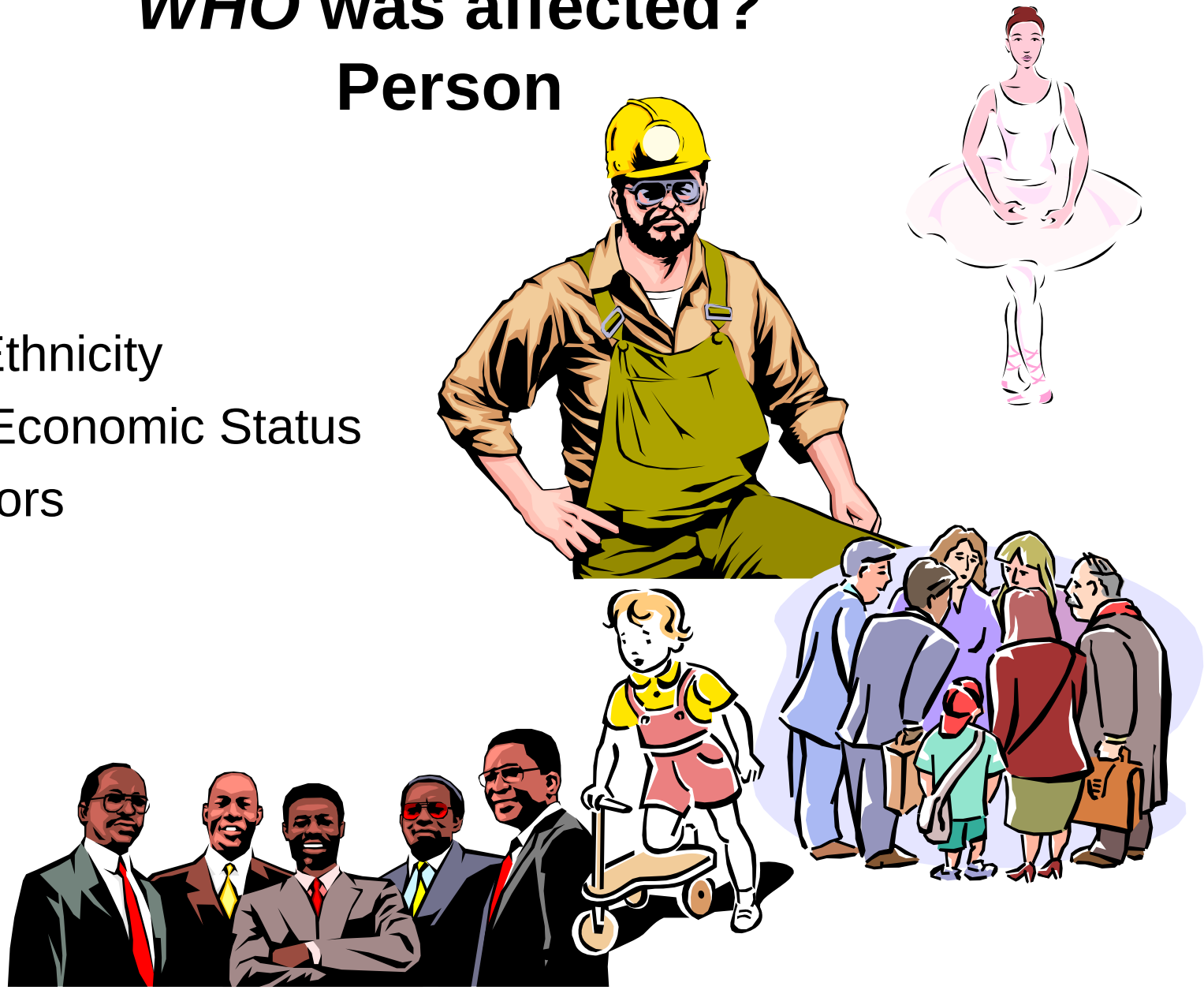
***WHERE* were they affected?**

***WHEN* were they affected?**

***HOW* and *WHY*?**

WHO was affected?
Person

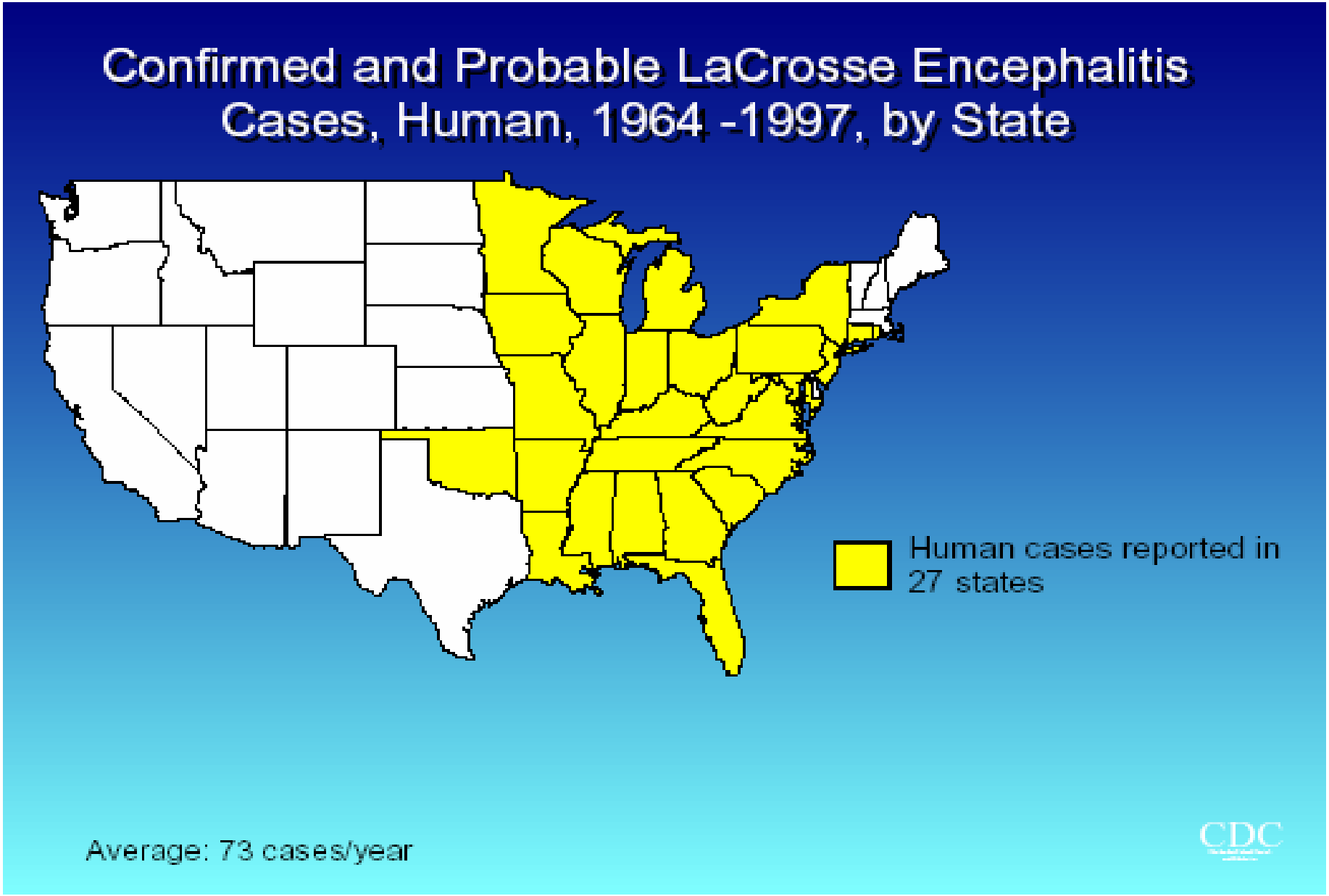
- Age
- Sex
- Race/Ethnicity
- Socio-Economic Status
- Behaviors

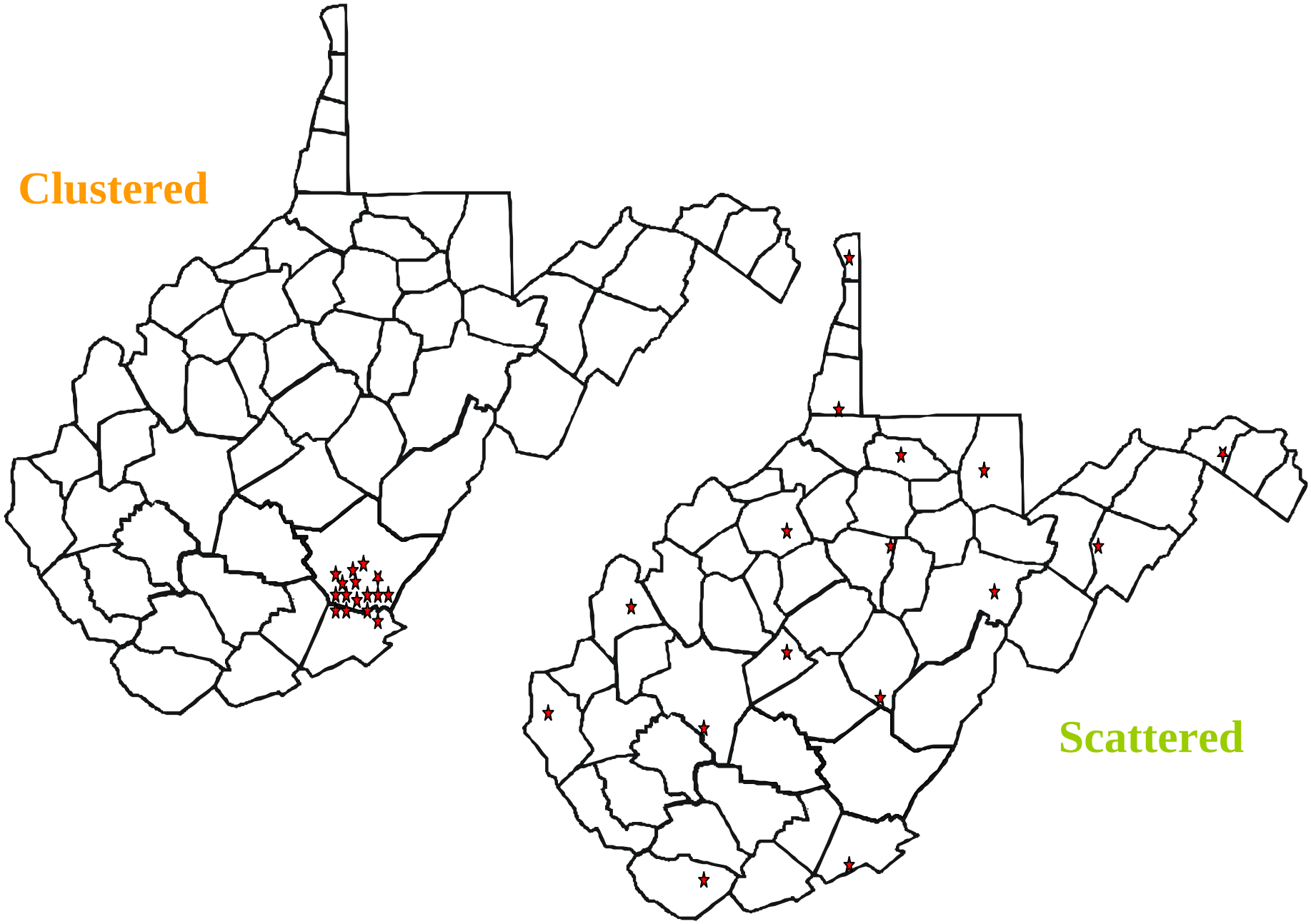


WHERE were they affected?
Place

- Geographic Distribution
  - Clustering, uniform, scattered
- Home
- Work
- School
- Hospital room

Geographic Distribution of LaCrosse

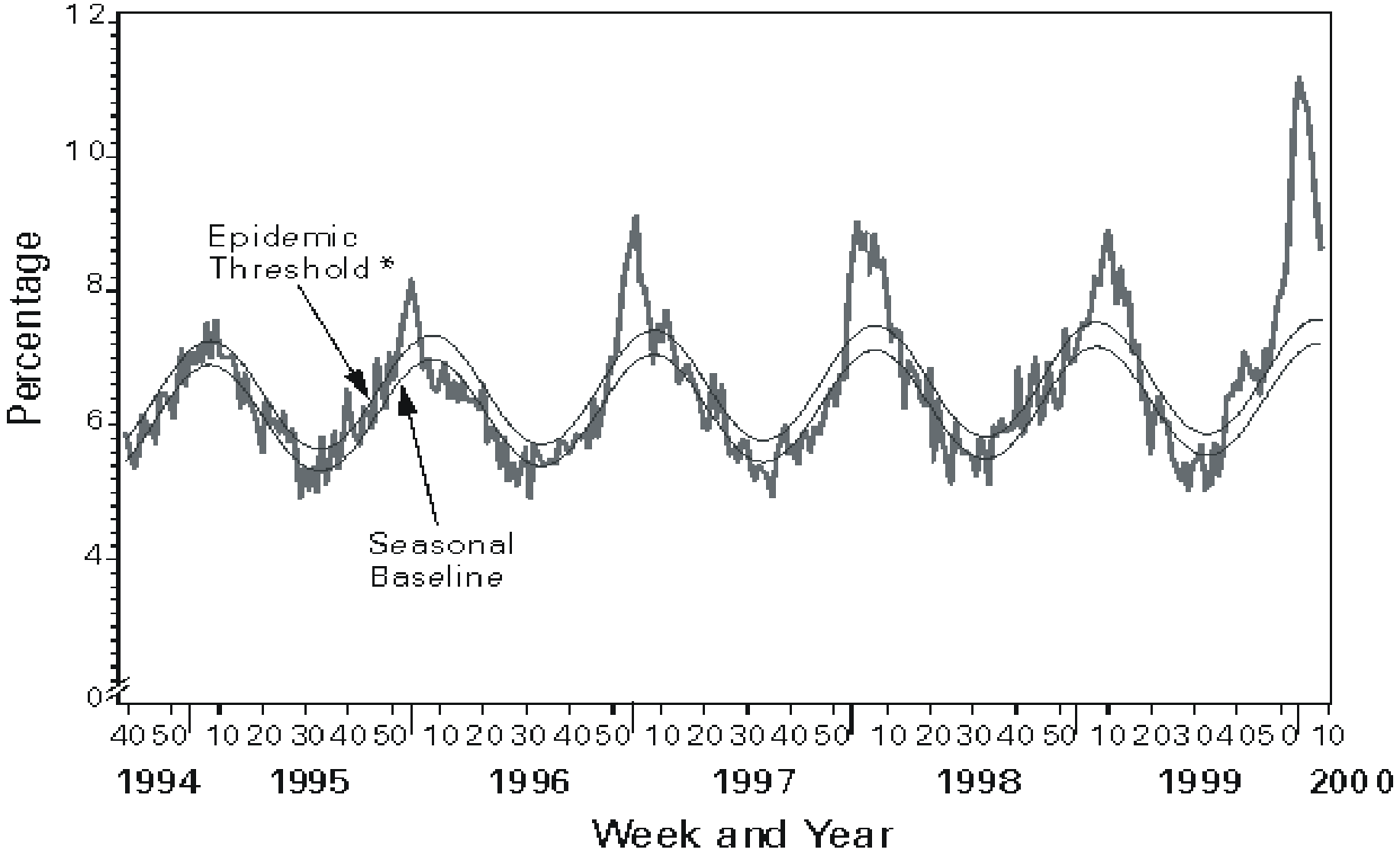



Clustered
Scattered

WHEN were they affected?
Time

- Onset of symptoms
- Incubation Period
- Infectious Period
- Seasonality
- Epidemic
- Interval
    - Long-term trends
    - Shorter for environmental exposure

FIGURE 1. Percentage of mortality attributable to pneumonia and influenza (P&I) in 122 cities, by week of report — United States, 1994–2000



*The epidemic threshold is 1.645 standard deviations above the seasonal baseline. The expected seasonal baseline is projected using a robust regression procedure in which a periodic regression model is applied to observed percentages of deaths from P&I since 1983.

Epidemiology key words: **Health-related states**

- Epidemiology is applied to the whole spectrum of health-related events:
  - chronic diseases,
  - environmental problems,
  - behavioral problems,
  - injuries,
  - infectious diseases.

Epidemiology key words: **Populations**

- One of the most important distinguishing characteristics of epidemiology is that it deals with groups of people rather than with individual patients.

Epidemiology key words: **Control**

- Epidemiology can be used simply as an analytical tool for studying diseases and their determinants.

- Epidemiological data also steers public health decision making and aids in developing and evaluating interventions to control and prevent health problems.

How is Epidemiology Used?

- Population/community health assessment
- Personal decision-making
- Complete clinical picture
- Evaluate interventions
- Search for cause
  - Exposure and relationship to disease
  - Outbreak investigation

Establishing a cause-and-effect relationship: Five criteria

- **Strength of association**—the relationship must be clear.
- **Consistency**—observation of the association must be repeatable in different populations at different times.
- **Temporality**—the cause must precede the effect.
- **Plausibility**—the explanation must make sense biologically.
- **Biological gradient**—there must be a dose-response relationship

Incidence

- Number of *new* events occurring in a defined population during a specified period of time
- Incidence =
  - New cases/population at risk/time
- Used to measure current disease activity
- Allows comparison between areas with different populations

Contrasts with Prevalence

- Prevalence is the number of new and existing cases divided by the total population (can be during a period of time or at a given point)
- Prevalence =

(New cases + existing cases)/Total population

  - Can be expressed as a percent
  - Can give a picture of disease burden within a population

Aim of epidemiological studies

- To determine distribution of disease
- To examine determinants of a disease
- To judge whether a given exposure causes or prevents disease

Exposure Variable – “E”

- Characteristic of interest
- Risk factor
- Predictor variable
- Independent variable
- Putative causal factor

Outcome or Disease Variable – “D”

- Health event of interest
- Illness, injury, infection
- Response variable
- Dependent variable
- Effect variable

Confounder variable

- A confounding factor in a study is a variable which is related to one or more of the variables defined in a study.

- A confounding factor may mask an actual association or falsely demonstrate an apparent association between the study variables where no real association between them exists.

- If confounding factors are not measured and considered, bias may result in the conclusion of the study.

Confounder variable (example)

- Early studies of the effects of alcohol on heart disease were confounded by smoking.
- Alcohol actually prevents heart disease; while smoking causes heart disease.
- Unfortunately for investigators, smoking and alcohol exposure are positively correlated (smokers drink more than non-smokers).
- Early studies appeared to show that alcohol caused heart disease; when actually the small protective effect of alcohol was masked by the large causative effect of smoking.
- Once smoking behavior was taken into account, the apparent negative effect (heart disease) of alcohol disappeared.

Definition of Bias

Bias is a systematic error that results in an incorrect (invalid) estimate of the measure of association

A. Bias can create spurious association when there really is none (bias away from the null)

B. Bias can mask an association when there really is one (bias towards the null)

C. Bias is primarily introduced by the investigator or study participants

Definition of Bias (cont'd)

D. Bias does not mean that the investigator is "prejudiced."

E. Bias can arise in all study types: experimental, cohort, case-control

F. Bias occurs in the design and conduct of a study. It can be evaluated but not fixed in the analysis phase.

G. Two main types of bias are selection and observation bias.

Selection Bias

A. Results from procedures used to select subjects into a study that lead to a result different from what would have been obtained from the entire population targeted for study

B. Most likely to occur in case-control or retrospective cohort because exposure and outcome have occurred at time of study selection

Selection Bias: What are the solutions?

- Little or nothing can be done to fix this bias once it has occurred.

- You need to avoid it when you design and conduct the study by, for example, using the same criteria for selecting cases and controls, obtaining all relevant subject records, obtaining high participation rates, and taking in account diagnostic and referral patterns of disease.

Observation Bias

- An error that arises from systematic differences in the way information on exposure or disease is obtained from the study groups
- Results in participants who are incorrectly classified as either exposed or unexposed or as diseased or not diseased
- Occurs after the subjects have entered the study
- Several types of observation bias: recall bias, interviewer bias, loss to follow up, and differential and non-differential misclassification

Observation Bias (Solutions)

- Use controls who are themselves sick; use standardized questionnaires that obtain complete information, mask subjects to study hypothesis

Categories of Epidemiology

- **Observational:** involve no intervention other than asking questions, carrying out medical examinations and simple laboratory tests or X-ray examinations:
  - **Descriptive**
  - **Analytical**

- **Experimental:** investigator intentionally alters one or more factors under controlled conditions to study the effects of doing so. The usual formal epidemiological experiment is the randomized control trial, which is done to test a preventive or therapeutic regimen or diagnostic procedure (Clinical trials).

Knowledge Continuum

- **Search for clues**
- **Clues available**

Descriptive Studies

- Patterns of occurrence
- No comparison group
- Generate hypotheses about E-D relationships

Analytic Studies

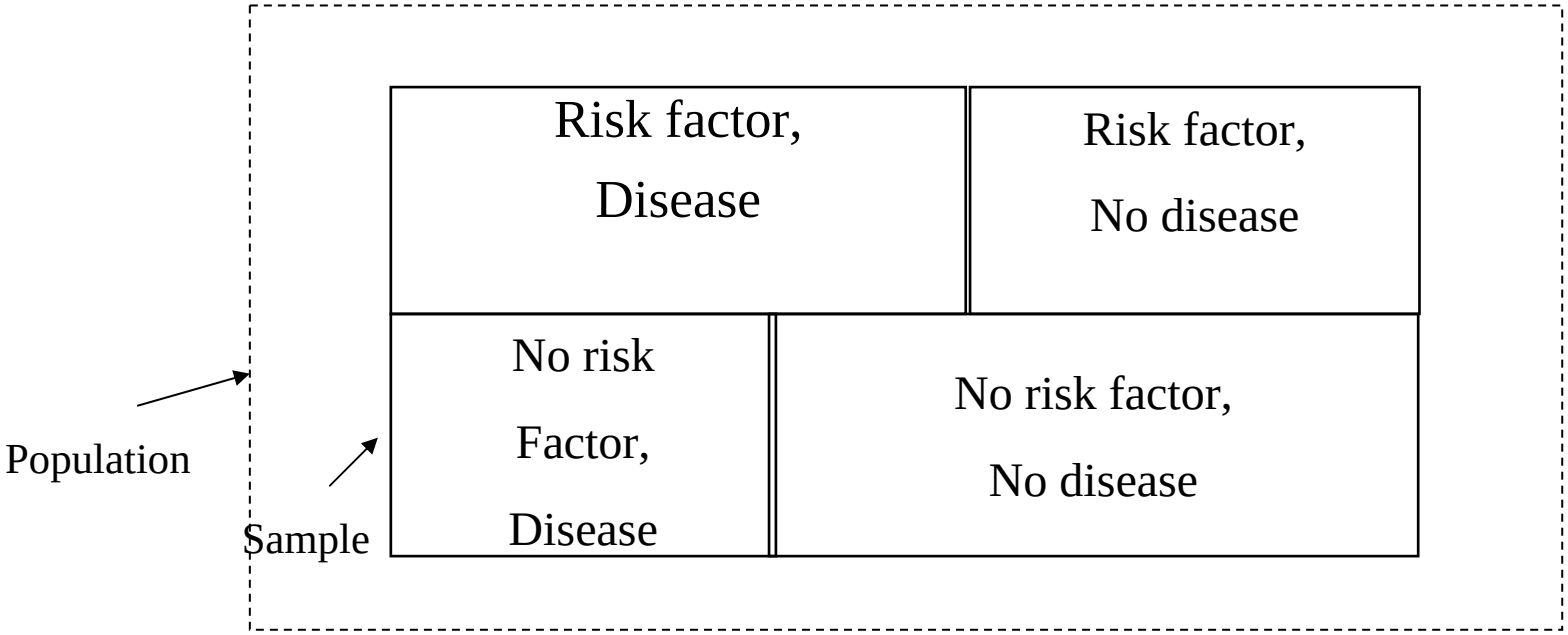
- Test hypotheses about E-D relationships

Types of Study Designs

- Cross-sectional
- Case-control
- Cohort:
  - Cohort means a group followed overtime
    - Two basic variations:
      - Prospective
      - Retrospective
- Experimental

Cross-sectional study

THE PRESENT



Steps:

1. Select a sample from the population
2. Measure predictor and outcome variables

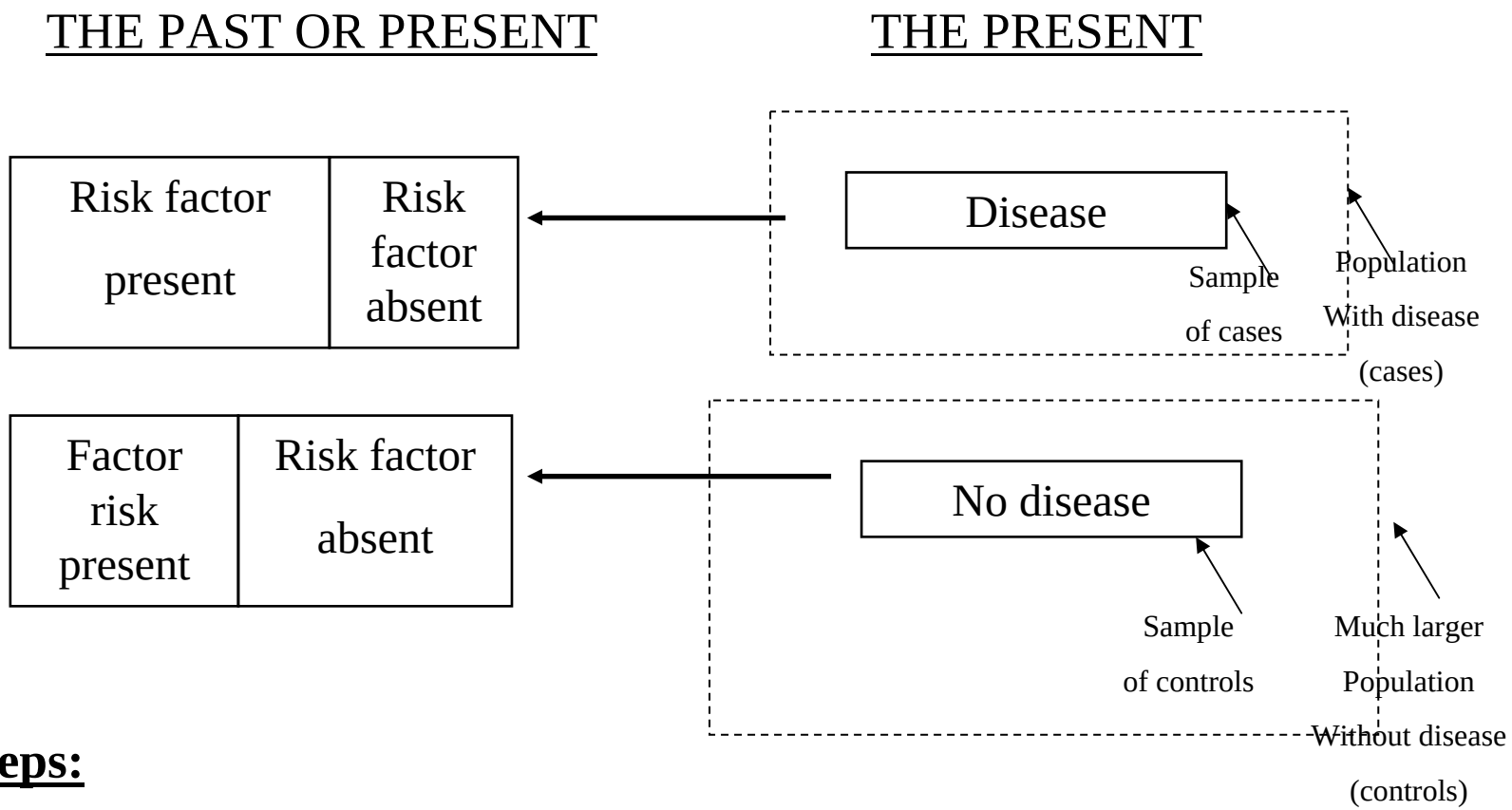
Major advantages & disadvantages of cross-sectional study

Advantages	Disadvantages
May study several outcomes	Not feasible for rare diseases
Short duration	Does not establish sequence of events
A good first step for a cohort study	Potential survivor bias
Yields prevalence	Does not yield incidence or true relative risk (measure of association)

Case-Control Studies

- Case-control studies are the most frequently undertaken analytical epidemiological studies
- They are the only practical approach for identifying risk factors for rare diseases
- They are best suited to the study of diseases for which medical care is sought, such as cancers or hip fracture

Case-control study



Steps:

1. Select a sample from the population of people with the disease (cases)
2. Select a sample from a population at risk that is free of the disease (controls)
3. Measure predictor variables

Selecting Cases

- Select cases after the diagnostic criteria and definition of the disease is clearly established
- Study cases should be representative of all cases
- The study need not include all cases in the population
- Population-based cases:
  - Include all subjects or a random sample of all subjects with the disease at a single point or during a given period of time in the defined population
- Hospital-based cases:

  All patients in a hospital department at a given time

Selecting Cases (cont.)

- Cases may be located from hospitals, clinics, disease registries, screenings, etc
- Incident cases are preferable to prevalent cases for reducing
  - (a) recall bias and
  - (b) over-representation of cases of long duration

- The most desirable way to obtain cases is to include all incident cases in a defined population over a specified period of time

Selecting Controls

- Controls should come from the same population at risk for the disease as the cases
- Controls should be representative of the target population
- Controls estimate the exposure rate to be expected in cases if there were no association between exposure and disease

Selecting Controls (cont.)

- Multiple controls can be used to help add statistical power when cases are unduly difficult to obtain
- Using more than one control group lends credibility to the results
- More than 3 controls for a case is usually not cost-efficient

Sources of cases and controls

CASES	CONTROLS
All cases diagnosed in the community	Sample of general population
All cases diagnosed in a sample of the population	Non-cases in a sample of the population
All cases diagnosed in all hospitals	Sample of patients in all hospitals who do not have the disease
All cases diagnosed in a single hospital	Sample of patients in the same hospital who do not have the disease
Any of the above methods	Spouses, siblings or associates of cases

Assessing Exposure

- Exposure is usually an estimate unless past measurements are available
  - It has to be assumed that the exposure incurred at the time the disease process began (this may not be valid)
- Exposure estimates are subject to *recall bias* and *interviewer bias*
  - Some protection may be afforded by blinding interviewers and carefully phrasing interview questions
- Potential confounders need to be accurately assessed in order to be controlled in the analysis

Possible Sources of Bias and Error

- Information on the potential risk factor (exposure) may not be available either from records or the study subjects' memories
- Information on potentially important confounding variables may not be available either from records or the study subjects' memories
- Cases may search for a cause for their disease and thereby be more likely to report an exposure than controls (recall bias)

Possible Sources of Bias and Error (cont.)

- The investigator may be unable to determine with certainty whether the suspected agent caused the disease or whether the occurrence of the disease caused the person to be exposed to the agent
- Identifying and assembling a case group representative of all cases may be unduly difficult
- Identifying and assembling an appropriate control group may be unduly difficult

Major advantages & disadvantages of case-control study

Advantages	Disadvantages
May study several exposures	Potential bias from sampling two populations
Short duration	Does not establish sequence of events
Relatively inexpensive	Potential bias in measuring predictors
Relatively small	Potential survivor bias
Yields odds ratio (measure of association)	Does not yield incidence or true relative risk (measure of association)

Odds Ratio (OR)

- A ratio that measures the odds of exposure for cases compared to controls
- Odds of exposure = number exposed ÷ number unexposed
- OR Numerator: Odds of exposure for cases
- OR Denominator: Odds of exposure for controls

Calculating the Odds Ratio

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

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

	OR<1	OR=1	OR>1
Odds comparison between cases and controls	Odds of exposure for cases are less than the odds of exposure for controls	Odds of exposure are equal among cases and controls	Odds of exposure for cases are greater than the odds of exposure for controls
Exposure as a risk factor for the disease?	Exposure reduces disease risk (Protective factor)	Particular exposure is not a risk factor	Exposure increases disease risk (Risk factor)

Interpreting the Odds Ratio

The odds of exposure for cases are 1.62 times the odds of exposure for controls.

or

Interpreting the Odds Ratio

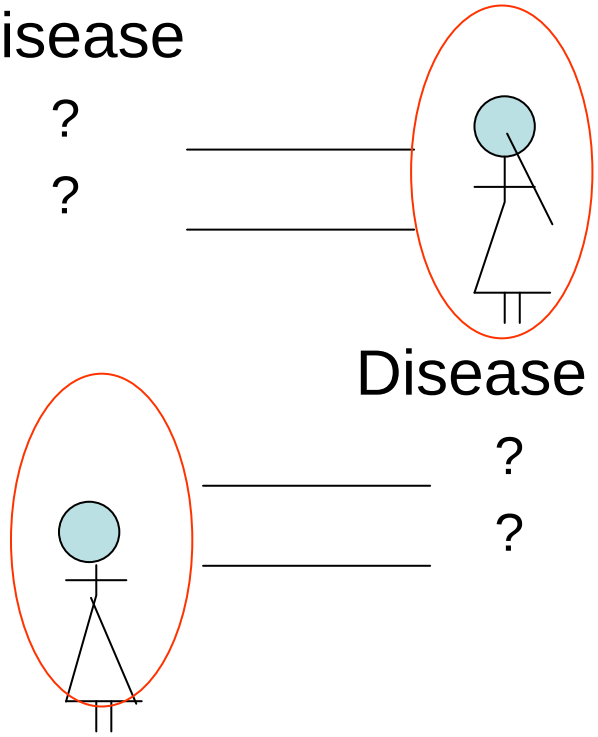
Those with CHD are 1.62 times more likely to be smokers than those without CHD

or

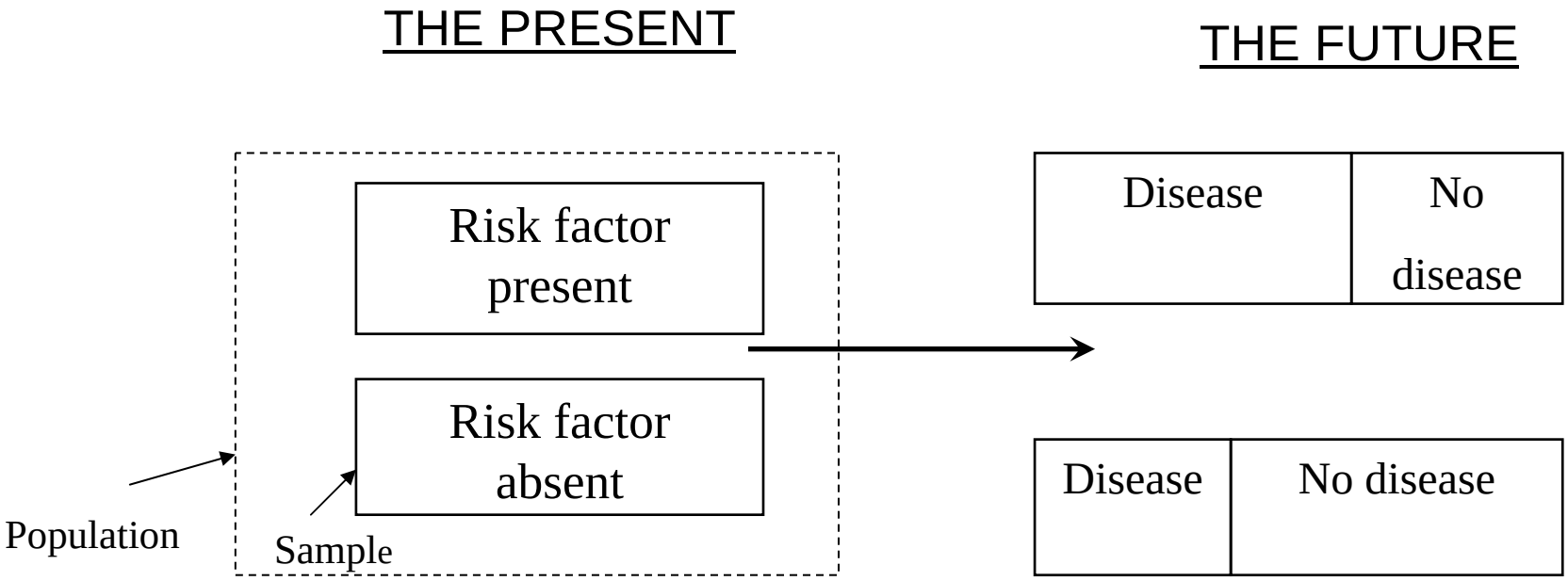
Those with CHD are 62% more likely to be smokers than those without CHD

Cohort studies

- Retrospective
  - Exposure — Disease
    - Yes — ?
    - No — ?
- Prospective
  - Exposure — Disease
    - Yes — ?
    - No — ?



Prospective cohort study



Steps:

1. Select a sample from the population
2. Measure predictor variables (risk factor present or absent)
3. Follow-up the cohort
4. Measure outcome variables (disease present or absent)

Selection of the Exposed Population

- Sample of the general population:
  - Geographically area, special age groups, birth cohorts (Framingham Study)
- A group that is easy to identify:
  - Nurses health study
- Special population (often occupational epidemiology):
  - Rare and special exposure
  - Permits the evaluation of rare outcomes

Selection of the Comparison Population

- **Internal Control Group**
  - Exposed and non-exposed in the same Study population (Framingham study, Nurses health study)
    - Minimise the differences between exposed and non-exposed
- **External Control Group**
  - Chosen in another group, another cohort (Occupational epidemiology: Asbestosis vs. cotton workers)
- **The General Population**

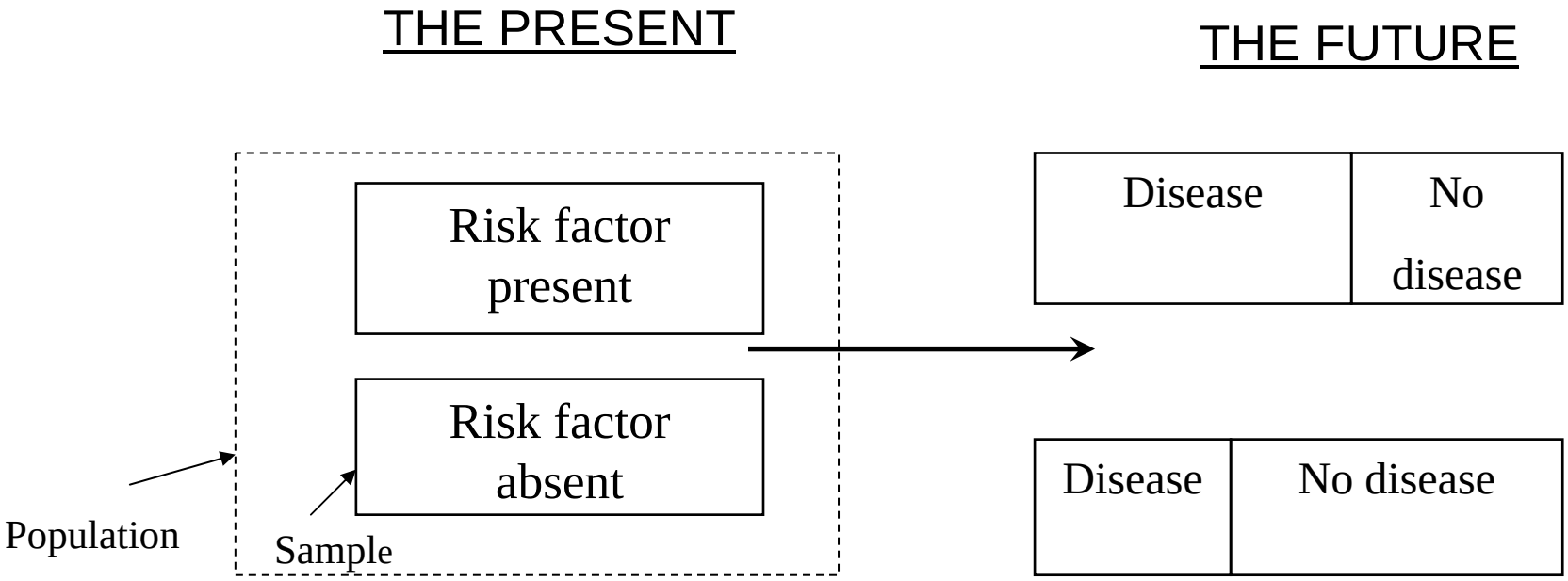
Bias in Cohort Studies

- **Selection bias:**
  - Non-response during data collection
  - Losses to follow up
  - Healthy worker effect
- **Misclassification on exposure or event**
  - Random
  - Systematic
- **Confounder**
  - Difference in other risk factors between exposed and non-exposed

Major advantages & disadvantages of cohort study

Advantages	Disadvantages
Establish sequence of events	Often requires large sample sizes
Avoids bias in measuring predictors	Not feasible for rare diseases
Avoids survivor bias	Expensive
Can study several outcomes	Long duration
More control over selection of subjects	
More control over measurements	
Yields incidence, relative risk (measure of association)	

Retrospective cohort study



Steps:

1. Identify a cohort assembled in past
2. Collect data on risk factors (measured in past)
3. Follow-up the cohort
4. Measure outcome variables (disease present or absent)

Major advantages & disadvantages of cohort study

Advantages	Disadvantages
Establish sequence of events	Often requires large sample sizes
Avoids bias in measuring predictors	Not feasible for rare diseases
Avoids survivor bias	Less control over selection of subjects
Relatively cheap	No control over nature and quality of measurements
Relatively shorter duration	Existing data may not include information that is important
Can study several outcomes	If data available, it may be: Incomplete or inaccurate
Yields incidence, relative risk (measure of association)	

Strengths in Cohort vs. Case-control?

Cohort study

- Rare exposure
- Examine multiple effects of a single exposure
- Minimizes bias in the in exposure determination
- Direct measurements of incidence of the disease

Case-control study

- Quick, inexpensive
- Well-suited to the evaluation of diseases with long latency period
- Rare diseases
- Examine multiple etiologic factors for a single disease

Limitations in Cohort vs. Case-control?

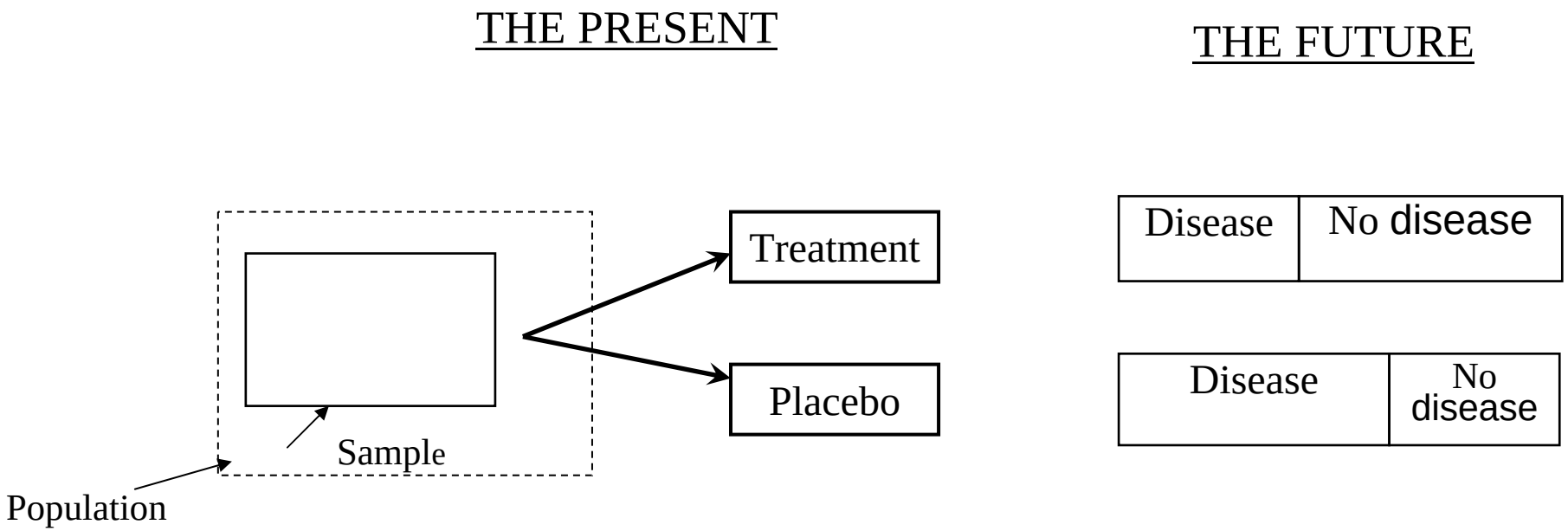
Cohort study

- Not rare diseases
- Prospective: Expensive and time consuming
- Retrospective: inadequate records
- Validity can be affected by losses to follow-up

Case-control study

- Not rare exposure
- Incidence rates cannot be estimated unless the study is population based
- Selection Bias and recall bias

Experimental study (Clinical Trials)



Steps:

1. Select a sample from the population
2. Measure baseline variables
3. Randomize
4. Apply interventions (one should be a blinded placebo, if possible)
5. Follow-up the cohort
6. Measure outcome variables (blindly, if possible)

Major advantages & disadvantages of experimental study

Advantages	Disadvantages
Can produce the strongest evidence for cause and effect	Costly in time and money
Only possible design for some research question (new drug)	Many research questions are not suitable: ethical barriers or outcome too rare
Can sometimes be faster and cheaper	Reduced generalizability (standardized interventions may be different than common practice)
	Tend to restrict the scope and narrow the study question