JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Bias and Confounding

Sukon Kanchanaraksa, PhD
Johns Hopkins University

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section A

Exposure and Disease Association

The Study Question

- An epidemiologic investigation ⇒ etiology of disease
  - Study hypothesis
    - A specific statement regarding the relationship between two variables: exposure and disease outcome

Association

- An epidemiologic study ⇒ test the hypothesis of association between exposure and outcome
  - If there is an association, the exposure is called a **risk factor** of the disease
- A risk factor can be either:
  - A **predictor** (marker or proxy)
    - Such as employment in a specific industry

    or
  - A **causal factor**
    - Such as exposure to benzene at work

From Association to Causation

- Steps in the study of the etiology of disease
- Limitations and issues in deriving inferences from epidemiologic studies
  - Bias and confounding
  - Criteria for causation
  - Interaction

Approaches for Studying the Etiology of Disease

- Animal models
- In-vitro systems
- Observations in human populations

Observations in Human Populations

Often begin with **clinical observations**

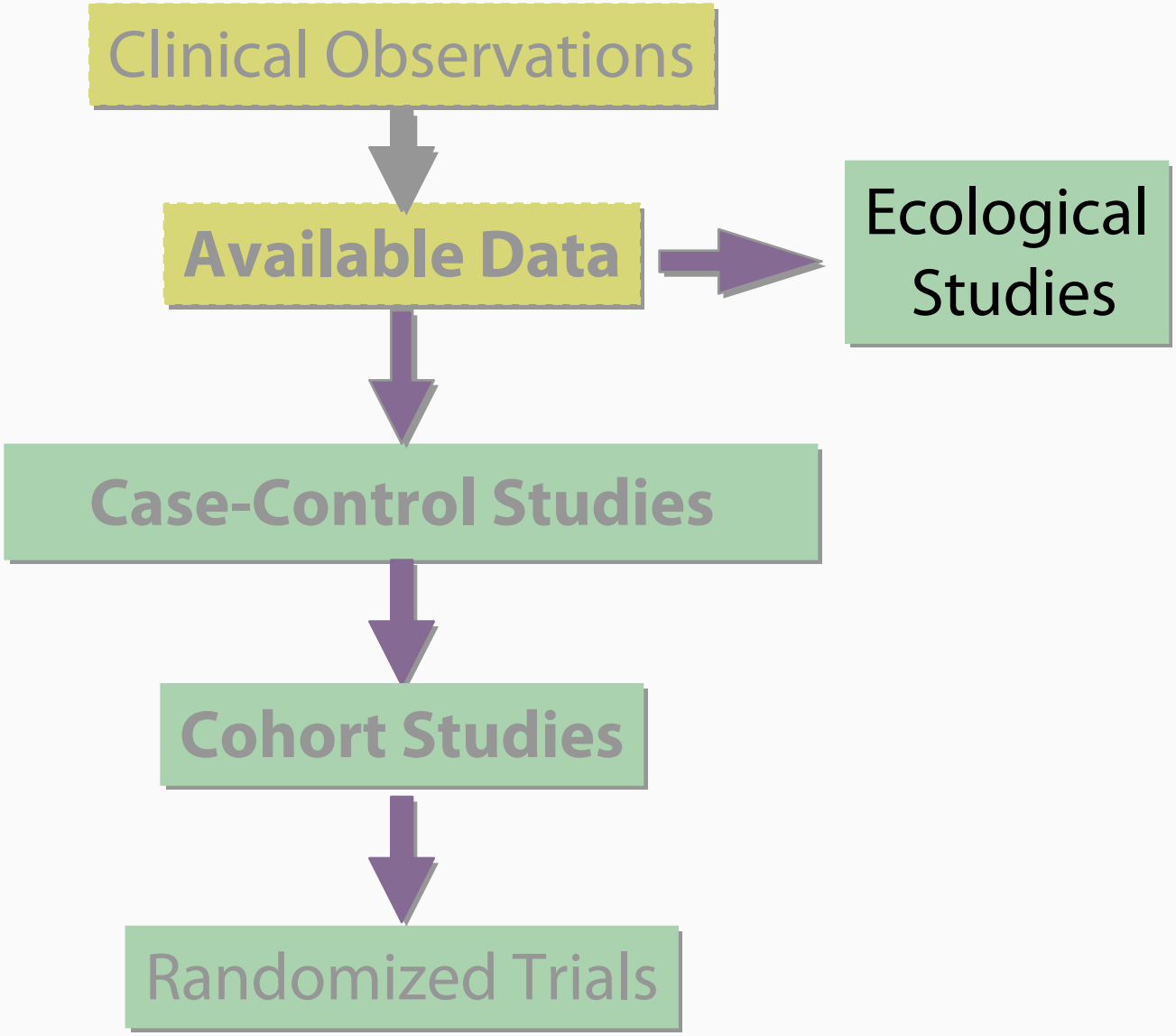
↓

Examine routinely **available data** to identify statistical associations

↓

Carry out new **studies** to demonstrate specific associations and derive causal inferences

Usual Sequence of Studies in Human Subjects

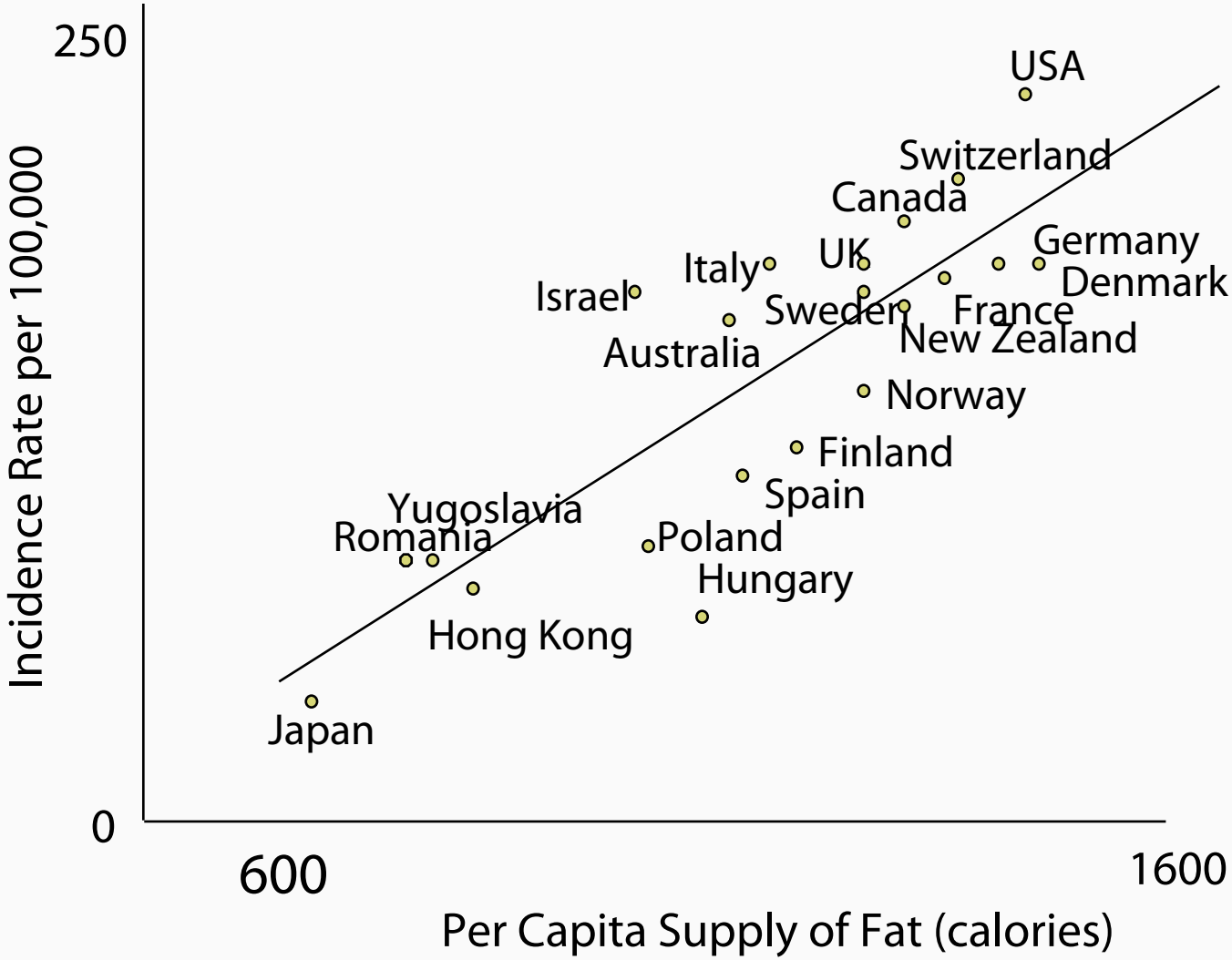


Ecological Study

- 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)

J Fagliano, M Berry, F Bove and T Burke. (1990). Drinking water contamination and the incidence of leukemia: an ecologic study. American Journal of Public Health, Vol. 80, Issue 10 1209-1212.
Grant WB. (1999). An ecologic study of dietary links to prostate cancer. Altern Med Rev. Jun;4(3):162-9.

Correlation between Dietary Fat Intake and Breast Cancer by Country

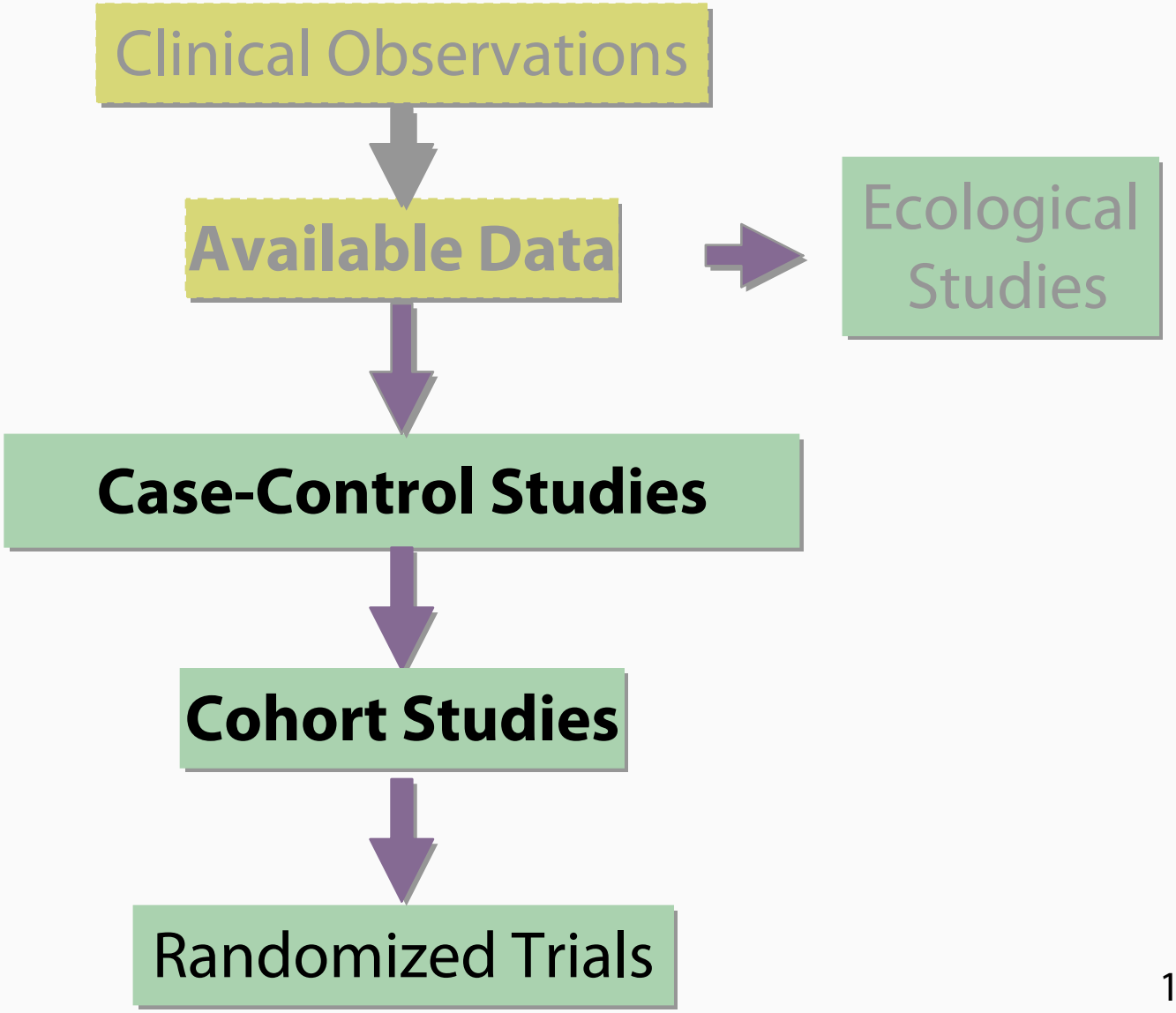


Source: Prentice, Kakar, Hursting, et al., Aspects of the rationale for the Womens' Health Trial. JNCI 80:802-814, 1988.

Ecological Fallacy

- **Ecological fallacy** is an error that could occur when an association between variables based on group (ecological) characteristics is used to make inferences about the association at an individual level when such association does not exist
- (On the contrary, **biological fallacy** is an error that may occur when the attempt to explain variations in population groups is based on individual study results)

Usual Sequence of Studies in Human Subjects



Observed Association

- If an association is observed, the first question asked must always be …

“Is it real?”

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section B

Bias and Confounding

Observed Association

- If an association is observed, the first question asked must always be …

“Is it real?”

Interpretation of Association

- Could it be by chance?
  - Chose a non-representative population to study (inadequate sample size)

- Could it be due to bias?
  - Bias is a systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease

    — (Schlesselman and Stolley, 1982)

Types of Bias

- Selection bias
- Information bias
- Confounding

Selection Bias

- **Selection bias** is a method of participant selection that distorts the exposure-outcome relationship from that present in the target population

A Case-Control Study of Alcoholism and Pneumonia

- Cases and controls selected from hospitalized patients
- Alcoholics with pneumonia are more likely to be admitted than non-alcoholics with pneumonia
- Risk of pneumonia associated with alcoholism is biased upwards

Pneumonia and Alcoholism in the Community

- In the **community**

Alcoholism \ Pneumonia	Yes	No
Yes	10	10
No	90	90
	100	100

$$OR = \frac{10 \times 90}{90 \times 10} = 1.0$$

Pneumonia and Alcoholism in the Hospital

- In the **hospital**

Alcoholism \ **Pneumonia**	Yes	No
Yes	20	10
No	80	90
	100	100

$$OR = \frac{20 \times 90}{80 \times 10} = 2.25$$

Selection Bias

- Selection bias occurs when the selection of participants in one group results in a different outcome than the selection for the other group

Examples of Selection Bias

- Select volunteers as exposed group and non-volunteers as non-exposed group in a study of screening effectiveness
  - Volunteers could be more health conscious than non-volunteers, thus resulting in less disease
  - Volunteers could also be at higher risk, such as having a family history of illness, thus resulting in more disease
- Study health of workers in a workplace exposed to some occupational exposures comparing to health of general population
  - Working individuals are likely to be healthier than general population that includes unemployed people (Healthy Worker Effect)
- Use prevalent cases instead of incidence cases

Controlling Selection Bias

- Define criteria of selection of diseased and non-diseased participants independent of exposures in a case-control study
- Define criteria of selection of exposed and non-exposed participants independent of disease outcomes in a cohort study
- Use randomized clinical trials

Information Bias

- **Information bias** occurs when information is collected differently between two groups, leading to an error in the conclusion of the association
- When information is incorrect, there is misclassification
  - **Differential misclassification** occurs when the level of misclassification differs between the two groups
  - **Non-differential misclassification** occurs when the level of misclassification does not differ between the two groups

Examples of Information Bias

- **Interviewer** knows the status of the subjects before the interview process
  - Interviewer may probe differently about exposures in the past if he or she knows the subjects as cases
- Subjects may **recall** past exposure better or in more detail if he or she has the disease (recall bias)
- **Surrogates**, such as relatives, provide exposure information for dead cases, but **living controls** provide exposure information themselves

Controlling Information Bias

- Have a standardized protocol for data collection
- Make sure sources and methods of data collection are similar for all study groups
- Make sure interviewers and study personnel are unaware of exposure/disease status
- Adapt a strategy to assess potential information bias

Confounding

- **Confounding** occurs when the observed result between exposure and disease differs from the truth because of the influence of the third variable
- For example, crude mortality rate (crude effect) of City A differs from the rate of City B—**but** after adjusting for age, the adjusted rates do not differ
  - Age distribution differs between the two cities
    - Age confounds the association

Bias and Confounding

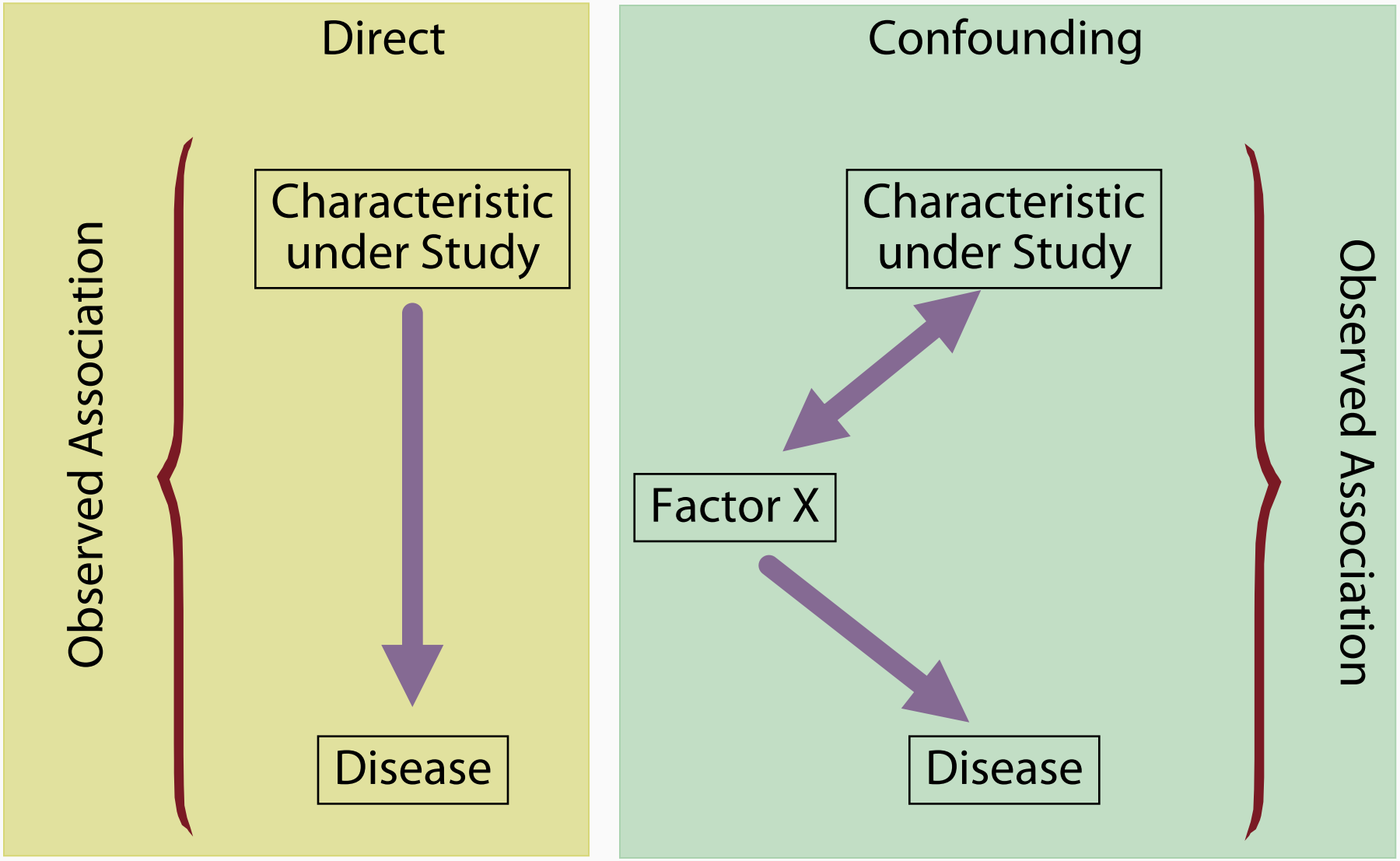
- **Bias** is a systematic error in a study and cannot be fixed
- Confounding may lead to errors in the conclusion of a study, but, when confounding variables are known, the effect may be fixed

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section C

Confounding

Types of Statistical Associations



Confounding

- Effect of a factor of interest is mingled with (confounded with) that of another factor
- Confounding is a situation in which a measure of the effect of an exposure is distorted because of the association of exposure with other factor(s) that influence the outcome under study
- Confounding occurs where an apparent association between a presumed exposure and an outcome is in fact accounted for by a third variable not in the postulated causal pathway; such a variable must be itself associated with both presumed exposure and outcome

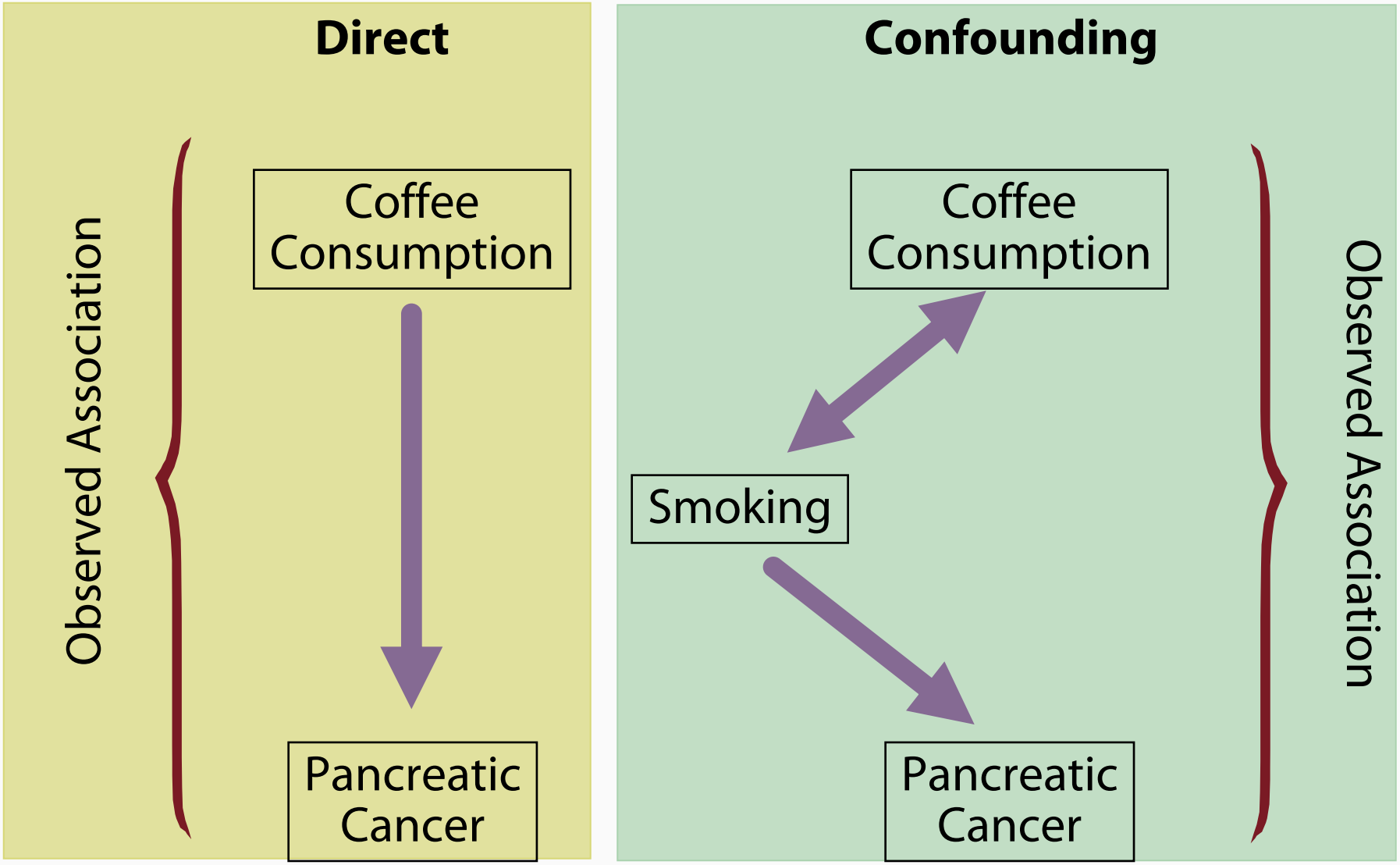
Confounding

- In a study of whether Factor A is a risk factor for Disease B, X is a **confounder** if:
  1. It is a risk factor for Disease B
  2. It is associated with Factor A (but is not a result of exposure to factor A)

Example of Confounding: Pancreatic Cancer Study

- In the study of whether coffee consumption is a risk factor for pancreatic cancer, smoking is a **confounder** if:
  1. It is a known risk factor for pancreatic cancer
  2. It is associated with coffee drinking but is not a result of coffee drinking

Types of Statistical Associations: Coffee Consumption and Pancreatic Cancer



Hypothetical Example of Confounding in a Case-Control Study

- In a study of 100 cases and 100 controls in an unmatched case-control study
  - 30% of cases and 18% of controls were exposed
  - OR was 1.95
- Could age confound the observed association?

Hypothetical Example of Confounding in an Unmatched Case-Control Study

Exposed	Cases	Controls
Yes	30	18
No	70	82
Total	**100**	**100**

$$OR = \frac{30 \times 82}{70 \times 18} = 1.95$$

Chi sq = 3.95

Observed association

Hypothetical Example of Confounding in an Unmatched Case-Control Study

- In order for age to be a confounder,
  1. Age must be a risk factor for the disease

     and
  2. Age must be associated with the exposure (but is not a result of the exposure)

Hypothetical Example of Confounding in an Unmatched Case-Control Study

Distribution of Cases and Controls by Age		
Age	Cases	Controls
< 40 years	50	80
• 4 0 years	50	20
Total	**100**	**100**

Chi sq = 19.8

Cases were older. So age meets criterion 1—age is a risk factor for the disease.

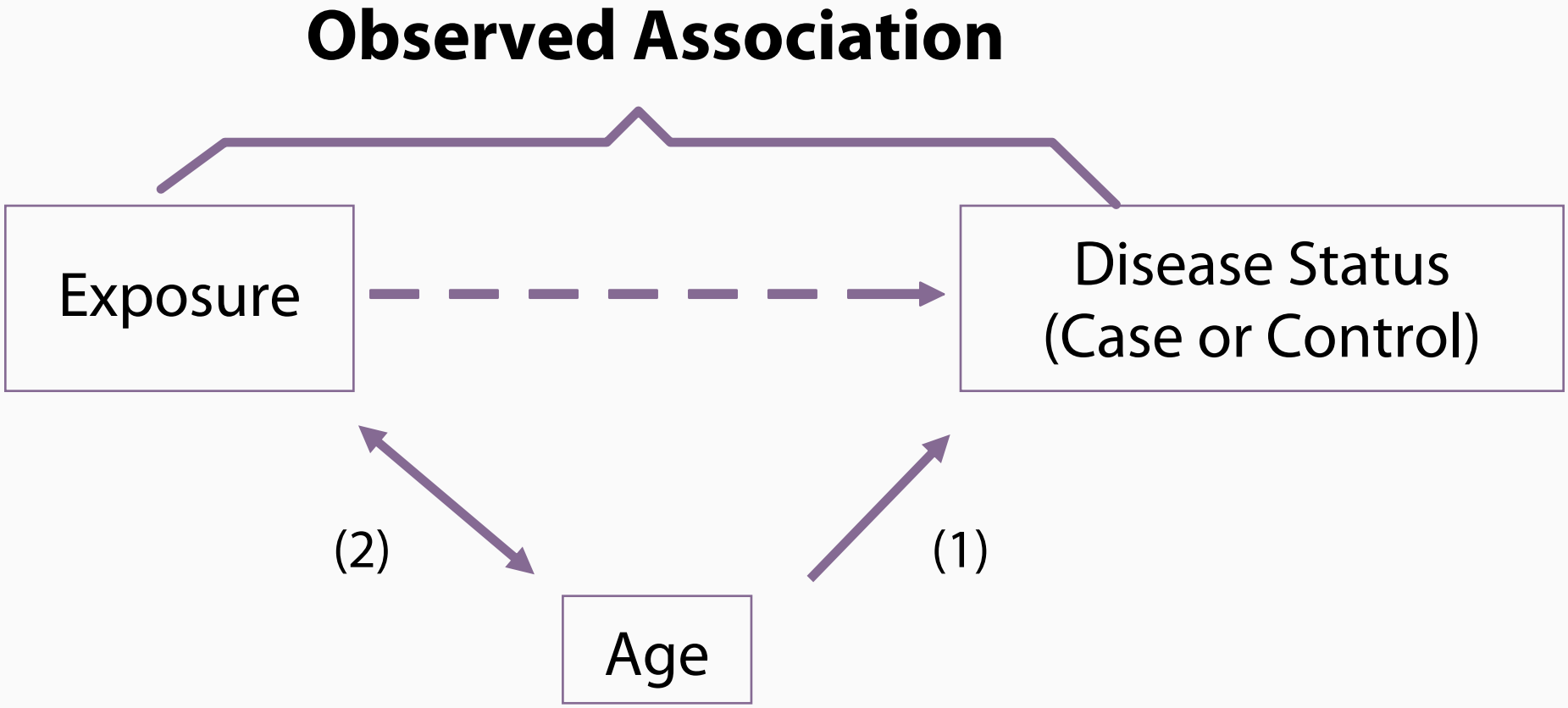
Hypothetical Example of Confounding in an Unmatched Case-Control Study

Relationship of Exposure to Age				
Age	Totals	Exposed	Not Exposed	Percent Exposed
< 40 years	130	13	117	10%
• 4 0 years	70	35	35	50%

Chi sq = 39.9

Older subjects were exposed more. So, age meets criterion 2—age is associated with exposure.

Hypothetical Example of Confounding in an Unmatched Case-Control Study



Therefore, age is a confounder

Hypothetical Example of Confounding in an Unmatched Case-Control Study

Calculations of Odds Ratios in a Stratified Analysis				
Age	Exposed	Cases	Controls	Odds Ratios
< 40 years	Yes	5	8	$\frac{5\times72}{45\times8}=\frac{360}{360}=1.0$
	No	45	72	
	Total	50	80	
• 4 0 years	Yes	25	10	$\frac{25\times10}{25\times10}=\frac{250}{250}=1.0$
	No	25	10	
	Total	50	20	

After stratified by age, observed association disappears

Hypothetical Example of Confounding in an Unmatched Case-Control Study II

Exposed	Cases	Controls
Yes	37	18
No	70	98
Total	**107**	**116**

$$OR = \frac{37 \times 98}{70 \times 18} = 2.9$$

Chi sq = 10.9

Observed association

Hypothetical Example of Confounding in an Unmatched Case-Control Study II

Calculations of Odds Ratios in a Stratified Analysis				
Age	Exposed	Cases	Controls	Odds Ratios
< 40 years	Yes	9	8	$\frac{9\times80}{45\times8}=\frac{720}{360}=2.0$
	No	45	80	
	Total	54	88	
• 4 0 years	Yes	28	10	$\frac{28\times18}{25\times10}=\frac{504}{250}=2.0$
	No	25	18	
	Total	53	28	

Age met both criteria for confounding. In this example, stratified ORs are not equal to 1.0. Age is a confounder.

Approaches to the Problem of Confounding

- In designing and carrying out the study
  - Matching
- In the data analysis
  - Stratification
  - Adjustment

Estimated Relative Risks of Pancreatic Cancer by Coffee-Drinking and Cigarette-Smoking

Estimated Relative Risks of Pancreatic Cancer by Coffee-Drinking and Cigarette-Smoking				
Cigarette-Smoking	Coffee-Drinking (Cups per Day)			Total
	0	1–2	• 3	
Never	1.0	2.1	3.1	1.0
Ex-smokers	1.3	4.0	3.0	1.3
Current smokers	1.2	2.2	4.6	1.2
Total	1.0	1.8	2.7	

Source: MacMahon et al, 1981.

When as Association does Exist

- To conclude that an association between exposure and disease outcome exists:
  - The study must have adequate sample size
  - The study must be free of bias
  - The study must be adjusted for possible confounders
- We can the pursue the original objective of whether the exposure is the **causal factor** of the disease