Introduction to Clinical Epidemiology

Kristen A. Stafford, MPH, PhD

Special thanks to Dr. Mona Baumgarten for the use of her slides

Outline

- Measures of disease occurrence
- Measures of association

- Epidemiologic study designs
- Bias
- Confounding
- Causality

What is Epidemiology?

Epidemiology is the ***study*** *of the* ***distribution*** *and* ***determinants*** *of* ***health-related states or events*** *in* ***specified populations****, and the* ***application*** *of this study to the control of health problems*

Dictionary of epidemiology. 4th ed. 2001

What is Clinical Epidemiology?

Clinical Epidemiology is the application of principles of epidemiology to clinical medicine. While classical epidemiology is the study of the distribution and determinants of diseases in populations, clinical epidemiology is the application of the principles and methods of epidemiology to conduct, appraise or apply clinical research studies focusing on prevention, diagnosis, prognosis, and treatment of disease. Clinical Epidemiology is the basic science of Evidence-based Medicine.

Measures of Disease Occurrence

- Prevalence
- Cumulative incidence
- Incidence rate

Prevalence

$$\frac{\text{Number of diseased individuals in population at a specified time}}{\text{Total population at same specified time}}$$

Prevalence

- Is a proportion and therefore has no units
- Ranges from 0 to 1
- Numerator includes both new and ongoing cases of disease
- Represents a cross-sectional “snapshot” of the population

Prevalence

- Does not estimate risk of disease
- Is not useful for studies of risk factors
- Estimates burden of disease
- Is useful in planning of health services

Example

- 10,600 men age 50-59 were examined in 2002 as part of a large heart health study
- 842 men were found to have coronary heart disease
- Prevalence of CHD = $0.079 \approx 8\%$

Cumulative Incidence

$$\frac{\text{Number of new cases of disease during specified time period}}{\text{Number of individuals at risk of disease at start of time period}}$$

At Risk

Individuals are at risk of disease if they:

- Do not have the disease at the start of the follow-up period
- Are capable of developing the disease
  - Have the organ of interest

Cumulative Incidence

- Represents the probability that an individual will develop the disease over a specified time period
- Is a measure of disease risk
- Is a proportion and therefore has no units
- Ranges from 0 to 1

Cumulative Incidence

- Based on assumption that all at-risk individuals are followed until they develop the disease or the observation period ends
- Does not reflect effect of differing lengths of follow-up
- *Syn.:* Incidence proportion

Example

- 10,600 men age 50-59 were examined in 2002 as part of a large heart health study
- 842 men were found to have coronary heart disease
- During the period 2002 to 2007, 317 men developed CHD
- Five-year cumulative incidence of CHD = ?

9,758 at risk   Cumulative incidence ≈ 3%

Incidence Rate

$$\frac{\text{Number of new cases of disease during specified time period}}{\text{Person-time of observation among people at risk during same time period}}$$

Incidence Rate

- Average rate at which a disease develops in a population over a specified time period
- Is a true rate and has the units of time
- Ranges from 0 to infinity
- Accounts for differing lengths of follow-up
- *Syn.:* Incidence density, hazard rate

Person-Time

Sum, over all individuals, of time at risk until the event of interest, loss to follow-up, or the end of the study

Example

Subject	Years of Follow-up	Got Disease
A	2	N
B	2	Y
C	1	N
D	1	N
E	1	Y
F	3	N
G	1	Y
H	1	Y
I	1	N
J	2	Y

Example

$$\text{Incidence rate} = \frac{\text{5 cases of disease}}{\text{15 person-years}}$$

$$= \text{0.33 cases per person-year}$$

$$= \text{33 cases per 100 person-years}$$

Challenge: person-years is epidemiologic jargon

33 new cases per 100 persons per year

Example

Interpretation:

New cases of the disease appear at the rate of

- 0.33 cases per person per year *or*
- 33 cases per 100 persons per year *or*
- 33 cases per 100 person-years

Incidence and Prevalence

- Change in incidence reflects change in etiologic factors (risk factors or protective factors)

- Change in prevalence reflects change in incidence or duration or both

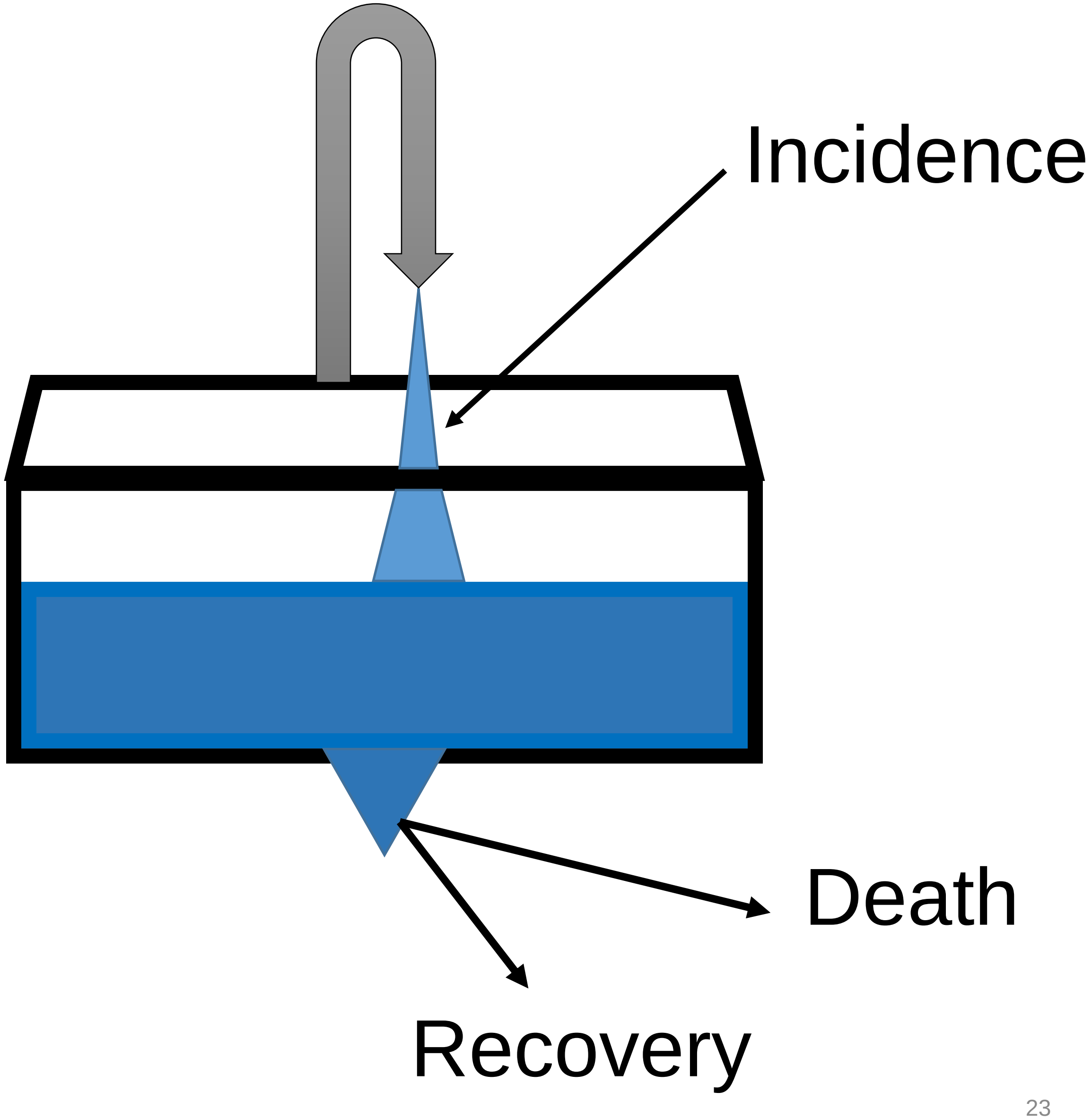
Incidence and Prevalence

$$\text{Prevalence} \approx$$
$$\text{incidence rate} \times \text{average duration of disease}$$

- Assumes incidence, prevalence, duration are stable over time
- Assumes prevalence < 10%

Prevalence

Not supposed to be
a fish tank



Measures of Association

- Relative risk

- Odds ratio

Relative Risk

- Ratio of disease incidence among exposed individuals to disease incidence among unexposed individuals
- Useful in research on disease etiology
- Quantifies magnitude of the association between an exposure and a disease
- *Syn.:* Risk ratio

Relative Risk

		CHD		
		Yes	No	
Smoking	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$$\text{Relative risk} = \frac{a/(a+b)}{c/(c+d)}$$

Relative Risk

- Varies from 0 to infinity
- When RR=1, there is no association between exposure and disease
- When RR > 1, the exposure is a risk factor for the disease, i.e., increases the risk of disease
- When RR < 1, the exposure is a protective factor for the disease, i.e., decreases the risk of disease

Example

		CHD Yes	CHD No	
Smoking	Yes	100	900	1,000
	No	20	980	1,000
		120	1,880	

Relative risk = ? 5

Odds

- Ratio of the probability that an event will occur to the probability that the event will not occur
- Risk = ratio of part to the whole
- Odds = ratio of part to the remainder
- Odds always higher than risk

Example

Rolling a die:

- Risk of rolling a 3 = 1/6 = 16.7%
- Odds of rolling a 3 = 1/5 = 20%

Example

- 20 smokers develop bronchitis while 30 do not
- Odds of bronchitis = ? 20/30 = 0.67
- Probability of bronchitis = ? 20/50 = 0.4

Odds Ratio

- Ratio of the odds of exposure among diseased to the odds of exposure among nondiseased

Odds Ratio

		CHD	
		Yes	No
Smoking	Yes	a	b
	No	c	d

$$\text{Odds ratio} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Odds Ratio

- Varies from 0 to infinity
- When OR=1, there is no association between exposure and disease
- When OR > 1, the exposure is a risk factor for the disease, i.e., increases the odds of disease
- When OR < 1, the exposure is a protective factor for the disease, i.e., decreases the odds of disease

Odds Ratio

- Only measure of association available from case-control studies
- Good estimate of the relative risk when the incidence is low (< 5% in the general population)

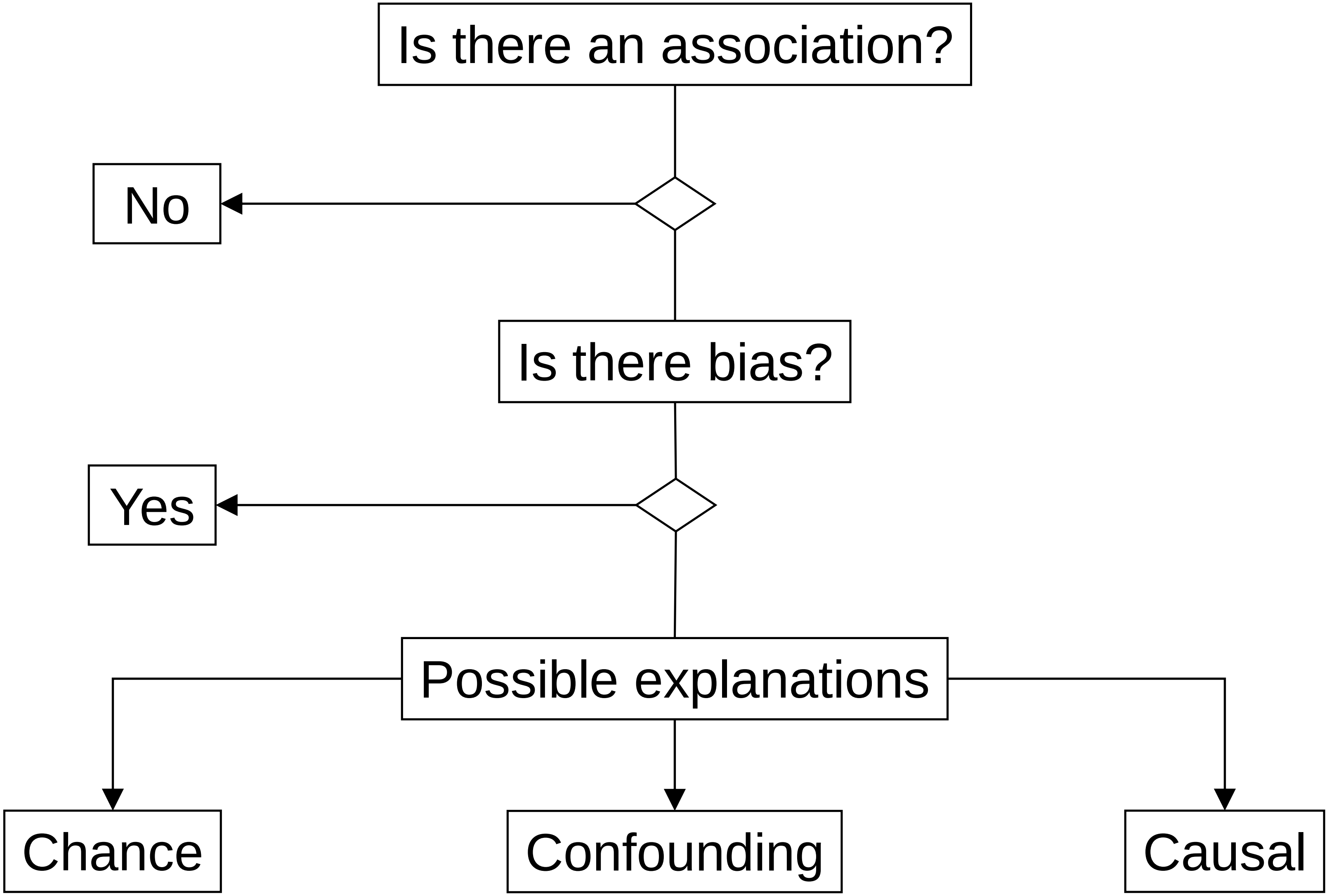
Example

		CHD Yes	CHD No
Smoking	Yes	25	10
	No	75	90
		100	100

Odds ratio = ? 3

Overview of Study Designs

Epidemiologic Reasoning



Epidemiologic Study Designs

Experimental Studies

- Randomized Controlled Trials
- Other Experimental Studies

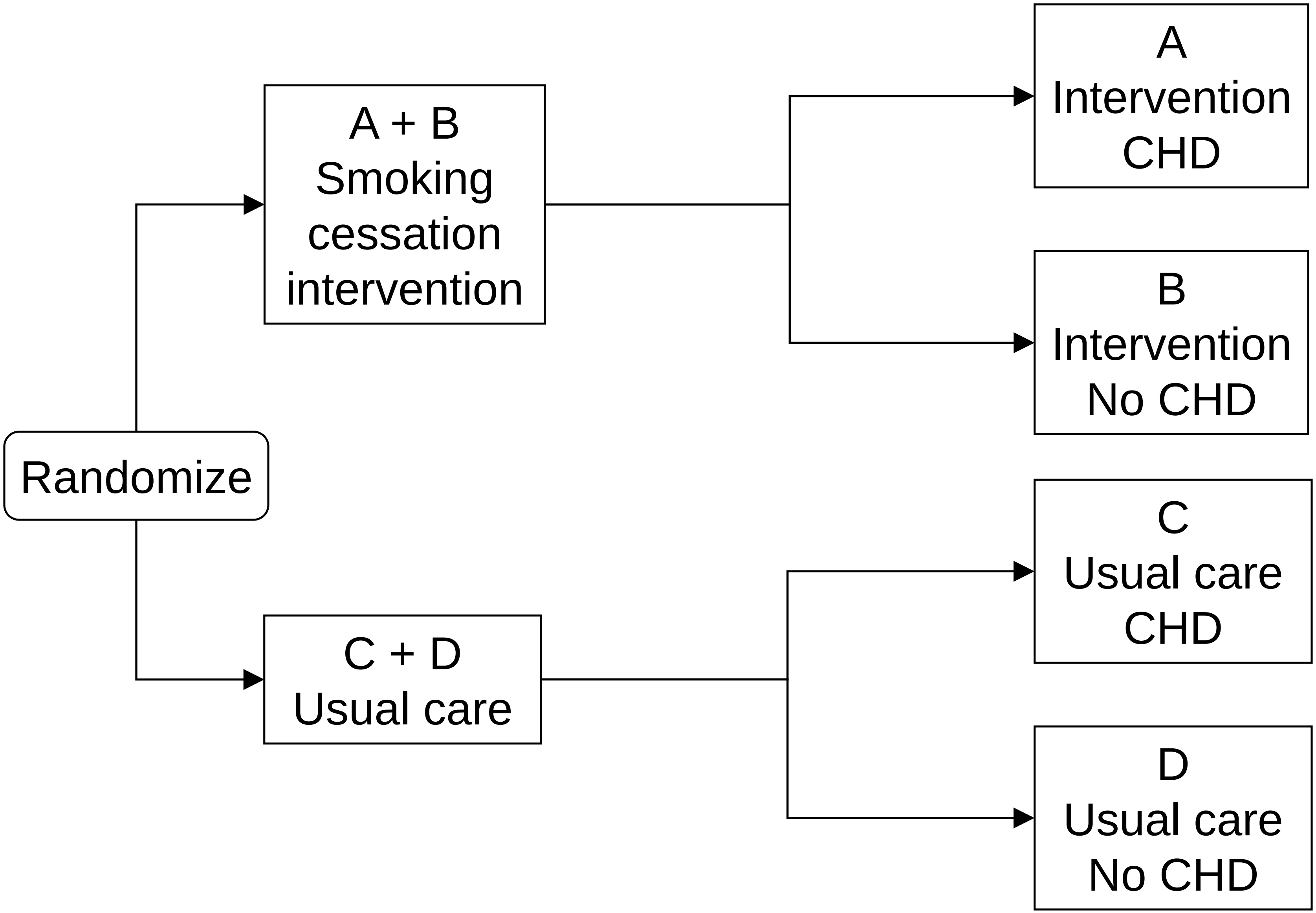
Observational Studies

- Cohort Studies
- Case-Control Studies
- Cross-Sectional Studies
- Ecologic Studies
- Case Series

Randomized Controlled Trials

- Treated and untreated subjects are followed over time to determine whether they experience the outcome (e.g., relapse, death, clinical improvement)
- Assignment to treatment or non-treatment is by randomization

Randomized Controlled Trials



Randomized Controlled Trials

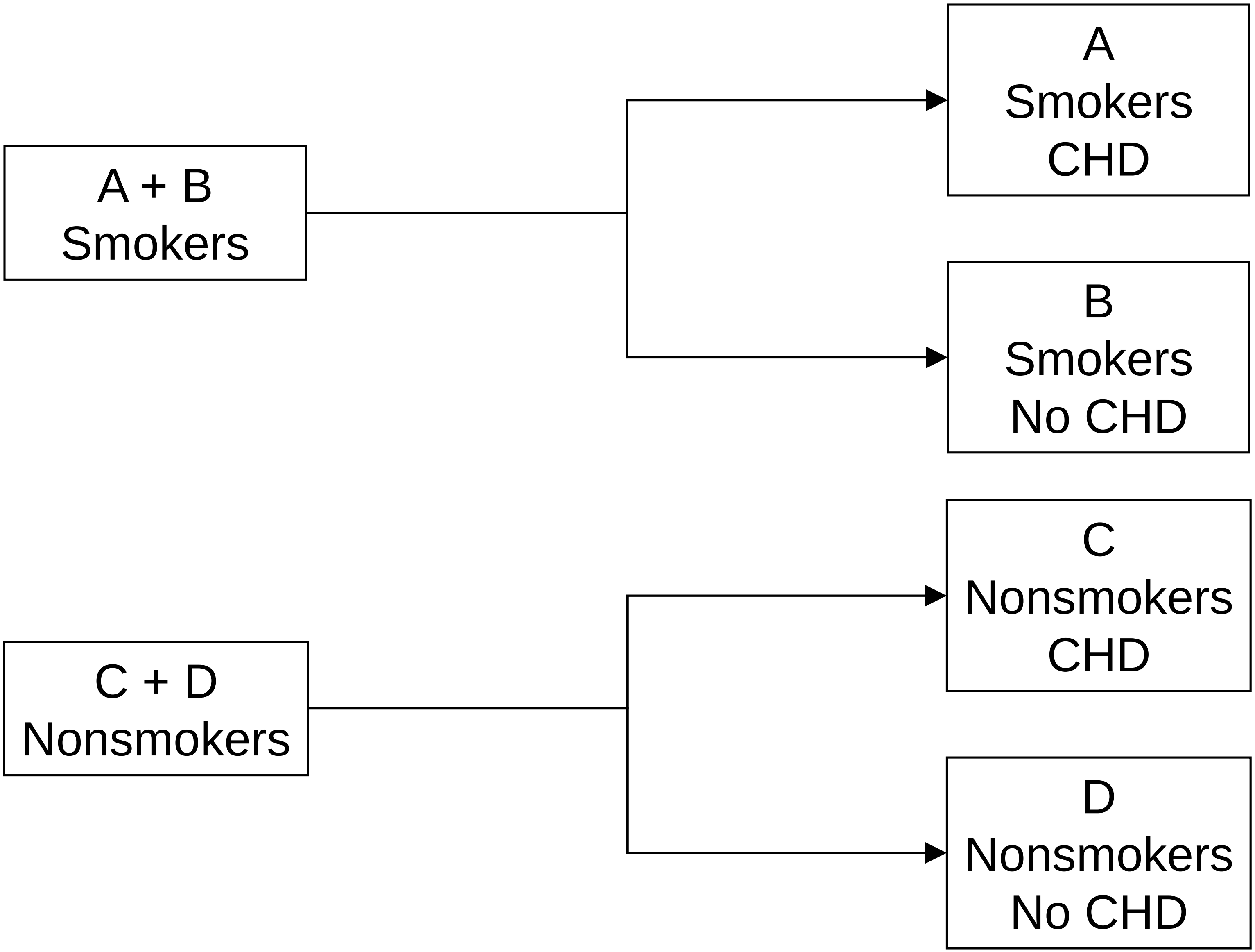
		CHD Yes	CHD No	
Intervention	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$$\text{Relative risk} = \frac{\text{CHD risk in intervention group}}{\text{CHD risk in usual care group}} = \frac{a/(a+b)}{c/(c+d)}$$

Cohort Studies

- Exposed and unexposed subjects without disease are followed over time to determine whether they experience the outcome
- Randomized controlled trials are a special case of the cohort study

Cohort Studies



Cohort Studies

		CHD Yes	CHD No	
Smoking	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$$\text{Relative risk} = \frac{\text{Risk of CHD among smokers}}{\text{Risk of CHD among nonsmokers}} = \frac{a/(a+b)}{c/(c+d)}$$

Case-Control Studies

- Compare exposure among persons with the disease (cases) to exposure among persons without the disease (controls)

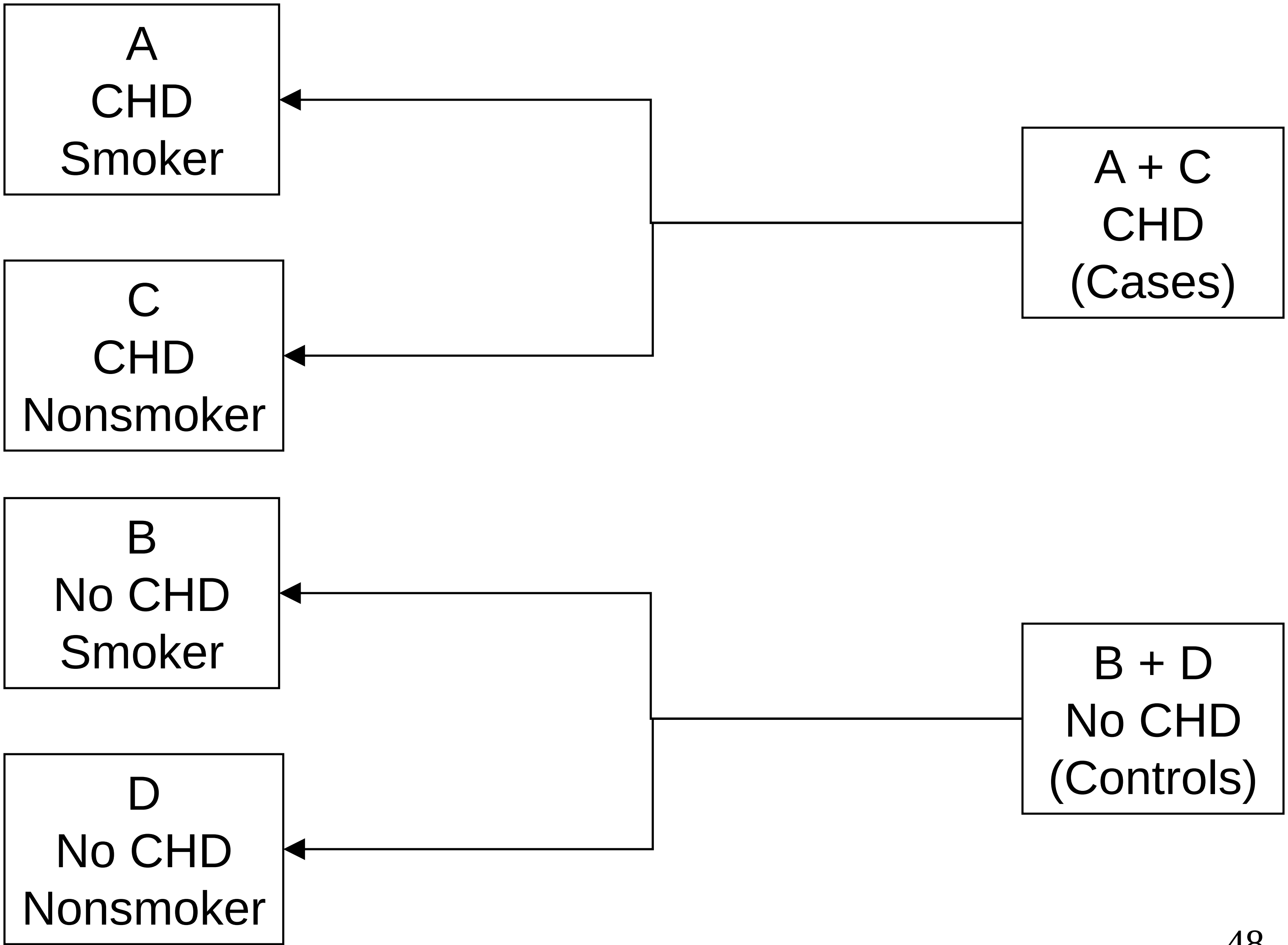
- Most commonly used epidemiologic study design despite many potential biases
  - If not designed well

- If designed well, can be thought of as an efficient cohort study
  - Measures of association can approximate rate ratios or risk ratios

Case-Control Studies

- More efficient than the equivalent cohort study
- Makes it possible to study rare diseases
- Makes it possible to study diseases that take a long time to develop
- Used for outbreak investigations

Case-Control Studies



Case-Control Studies

		CHD	
		Yes	No
Smoking	Yes	a	b
	No	c	d
		a+c	b+d

$$\text{Odds ratio} = \frac{\text{Odds of smoking among CHD cases}}{\text{Odds of smoking among controls}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Cross-Sectional Studies

- Study in which the status of individuals with respect to one or more characteristics is assessed at one point in time

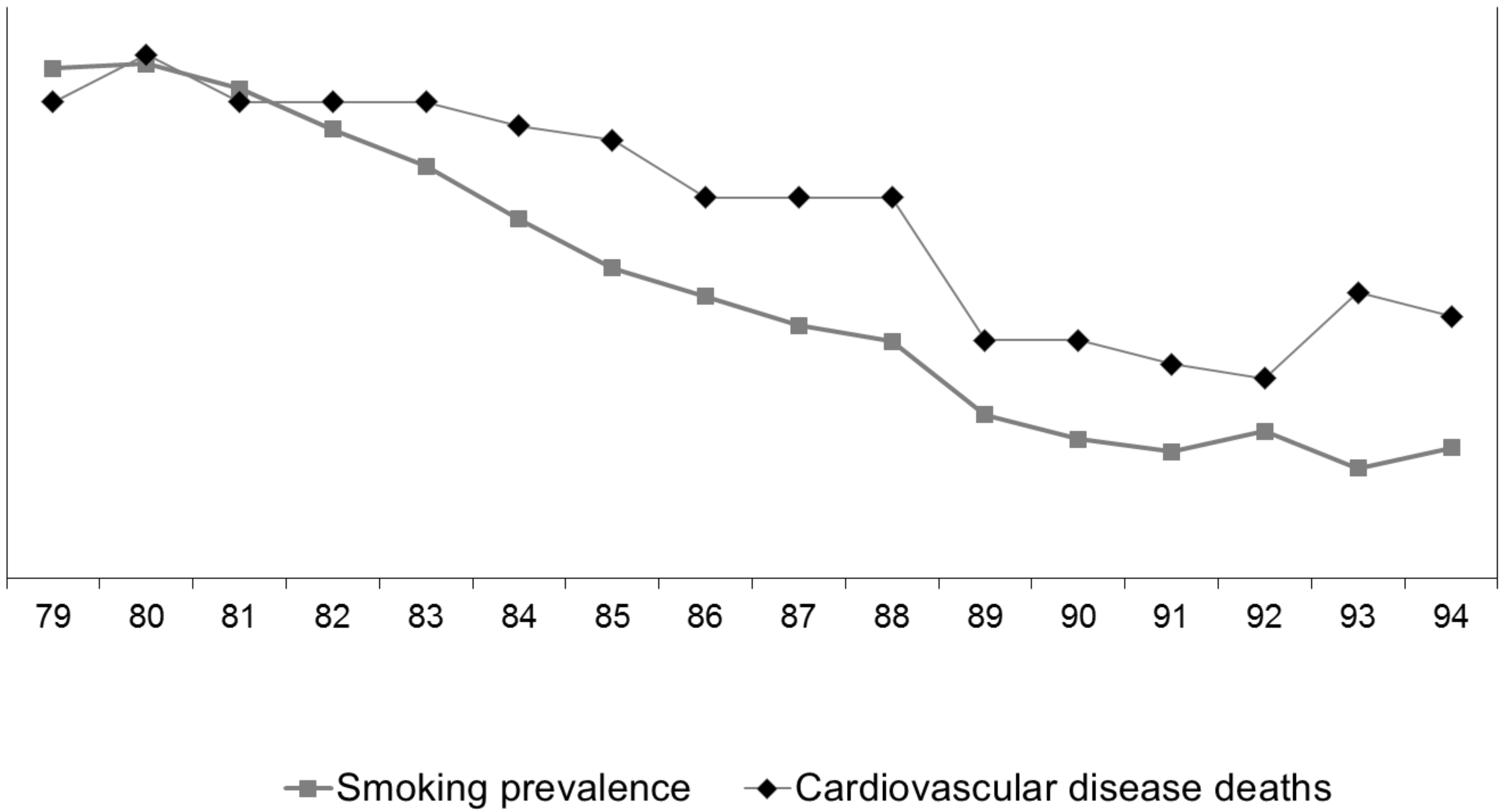
Cross-Sectional Studies

- May not be possible to determine whether exposure preceded disease
- No distinction between new cases and existing cases
- Not useful for the study of etiologic factors

Ecologic Studies

- Studies in which the units of analysis are populations or groups of people, rather than individuals
- Useful for hypothesis generation

Cardiovascular Disease Deaths and Smoking Prevalence (Males, 1979-1994)



Ecologic Fallacy

- Each individual in the population is characterized by the average for the population
- Bias may occur because an association observed between variables on an aggregate level does not necessarily represent the association that exists at an individual level
  - Because you don’t know the joint distribution of exposure/disease/other factors at an individual level

Case Series

- Studies without a comparison group
- All study subjects have the disease (or the exposure)
- Impossible to make inferences about causality
- Usually the first report of a new disease/syndrome
    - HIV, microcephaly due to Zika

Example

- 30% of a series of CHD patients are found to be smokers
- Can we conclude that there is an association between CHD and smoking?

Bias

- Deviation of results or inferences from the “truth”
- *Antonym:* Validity

Bias

- Selection bias
- Information bias

Selection Bias

- Distortion in study results due to the manner in which subjects are selected for the study

Examples of Selection Bias

- Bias related to nonresponse
- Bias related to loss to follow-up

Nonresponse

- Nonresponse may be due to refusal, migration, death, missing records
- Nonrespondents may differ from respondents

Nonresponse

Example:

- Subjects who refuse to participate in a study of smoking and CHD may be more likely to be smokers

Loss to Follow-Up

- In cohort studies and randomized controlled trials, persons who are lost to follow-up may differ from those who remain in the study

Loss to Follow-Up

Example:

- Prospective cohort study of the effect of smoking on CHD
- Study dropouts may be more likely to be smokers

What Can Be Done?

- Be aware of potential sources of selection bias
- Proper study design

Information Bias

- Errors in classification of subjects with respect to disease or exposure

Information Bias

Example:

- Case-control study of CHD and smoking
- Persons with CHD may be more likely to deny smoking history

What Can Be Done?

- Use data collection tools that have been validated, pretested
- Use similar data collection methods for all subjects in study (cases/controls, exposed/unexposed)
- Ensure that research staff is “blind” to subjects’ disease and exposure status

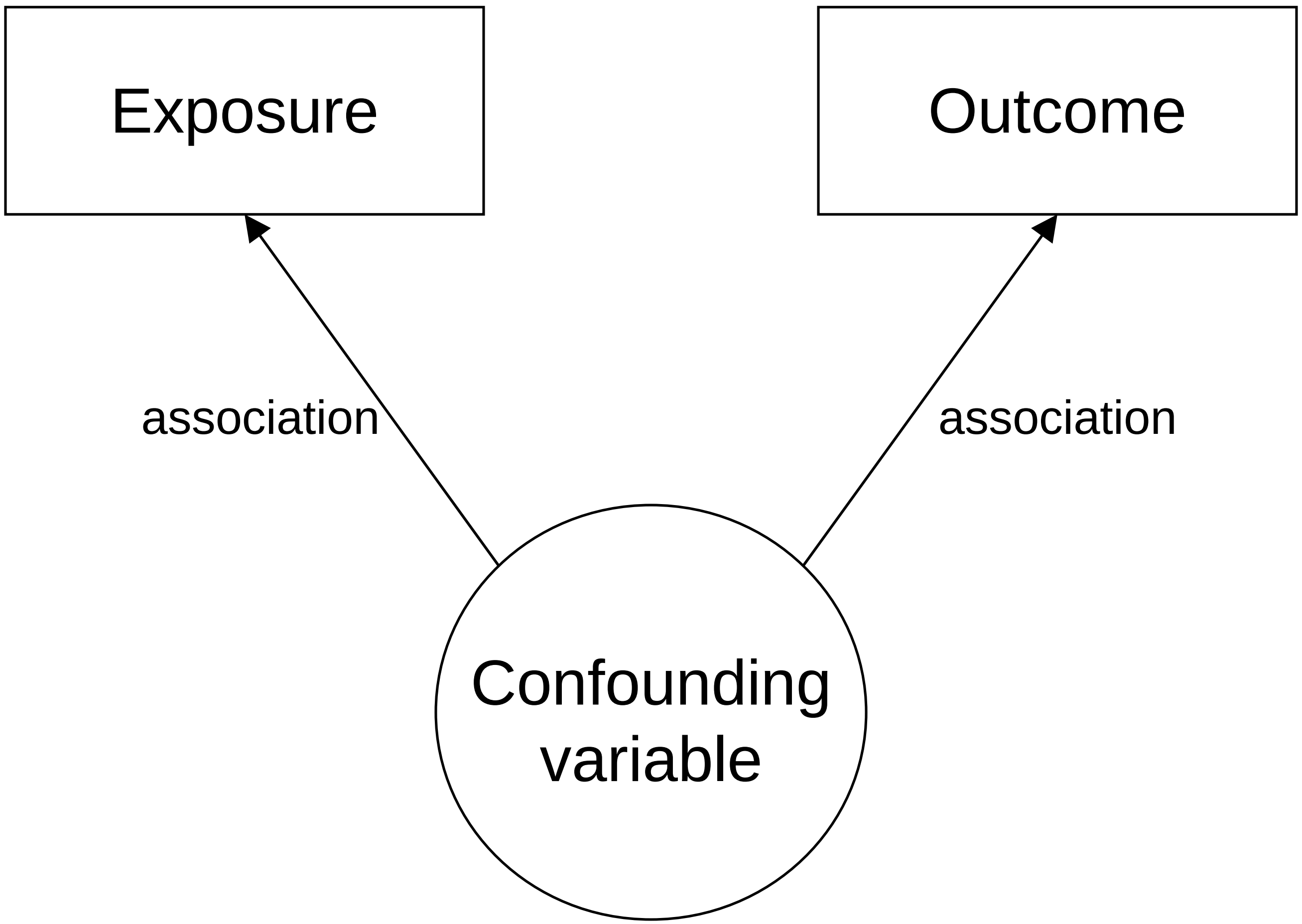
Confounding

- Confounding is the distortion of an exposure-outcome association brought about by the association of another factor with both outcome and exposure
- A confounder is a variable that masks the true relationship between an exposure and a disease

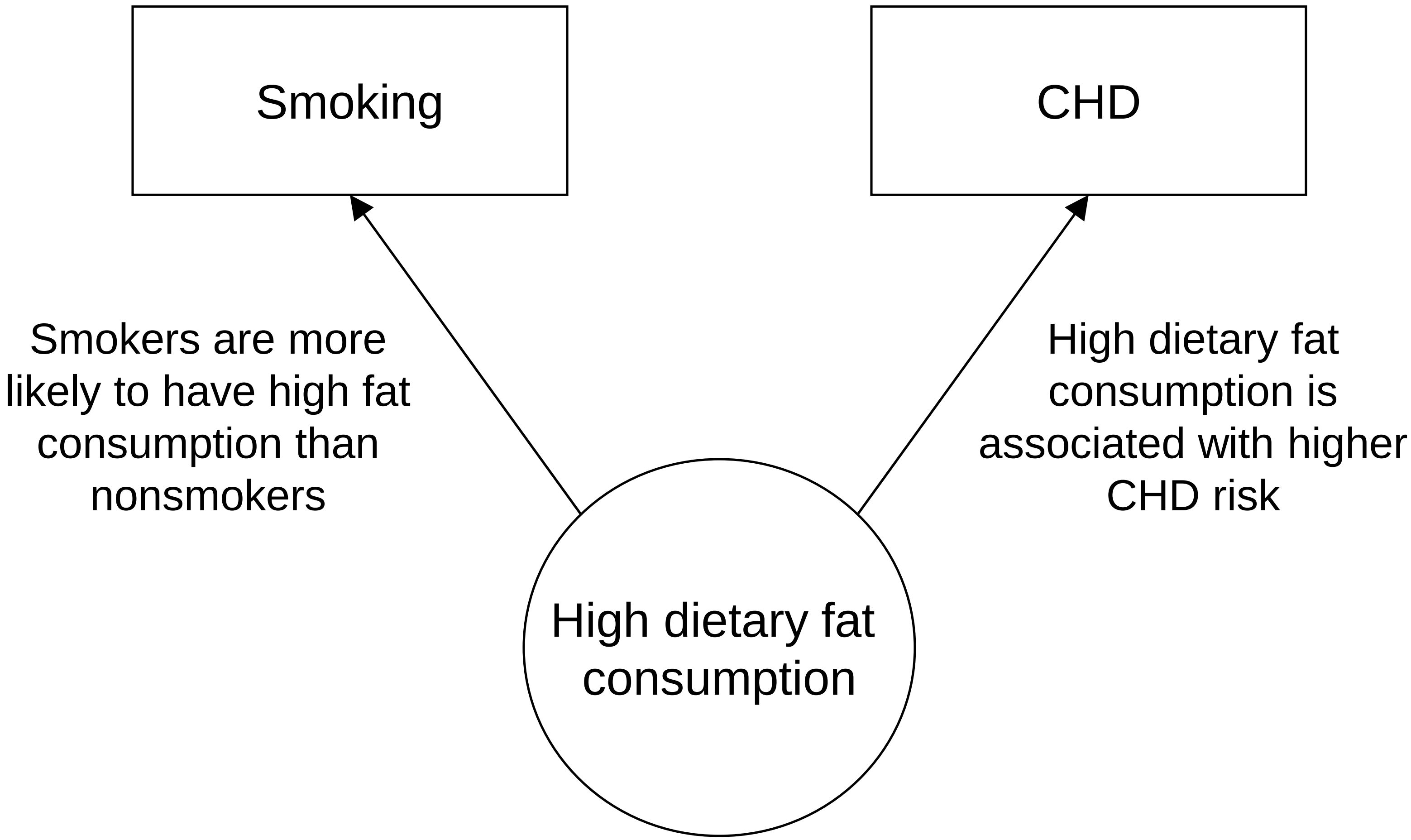
Confounding

- In order for confounding to occur, a variable must be a risk factor for the disease *and* be distributed differently among exposed and nonexposed
- If only one of these conditions is met, there will be no confounding

Confounding



Example

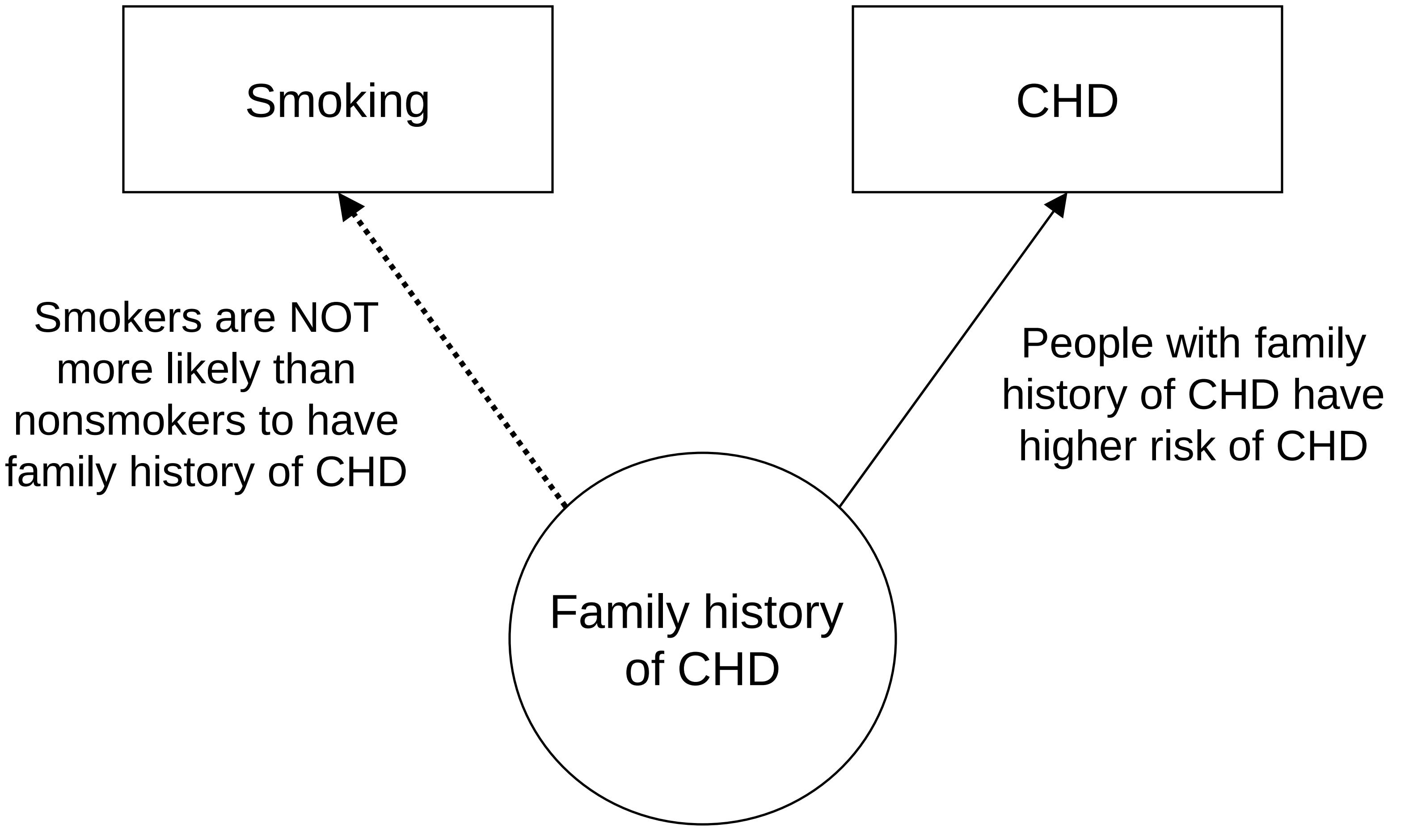


Smoking
CHD
Smokers are more likely to have high fat consumption than nonsmokers
High dietary fat consumption is associated with higher CHD risk
High dietary fat consumption

Example

- Suppose you wish to study the effect of smoking on the risk of CHD
- Smokers are more likely to have high dietary fat consumption than nonsmokers
- High dietary fat consumption is a risk factor for CHD
- Therefore, high dietary fat consumption is a confounder

Example



Smoking
CHD
Smokers are NOT more likely than nonsmokers to have family history of CHD
People with family history of CHD have higher risk of CHD
Family history of CHD

Example

- Suppose you wish to study the effect of smoking on the risk of CHD
- Family history of CHD is a risk factor for CHD
- Family history of CHD is not more common in smokers than nonsmokers
- Therefore, family history of CHD is not a confounder

Control of Confounding

- If a variable is a confounder, then controlling for that variable will result in a change in the estimated effect of the exposure on the disease

Control of Confounding

At design stage:

- Randomization
- Matching
- Restricting study to certain groups

At analysis stage:

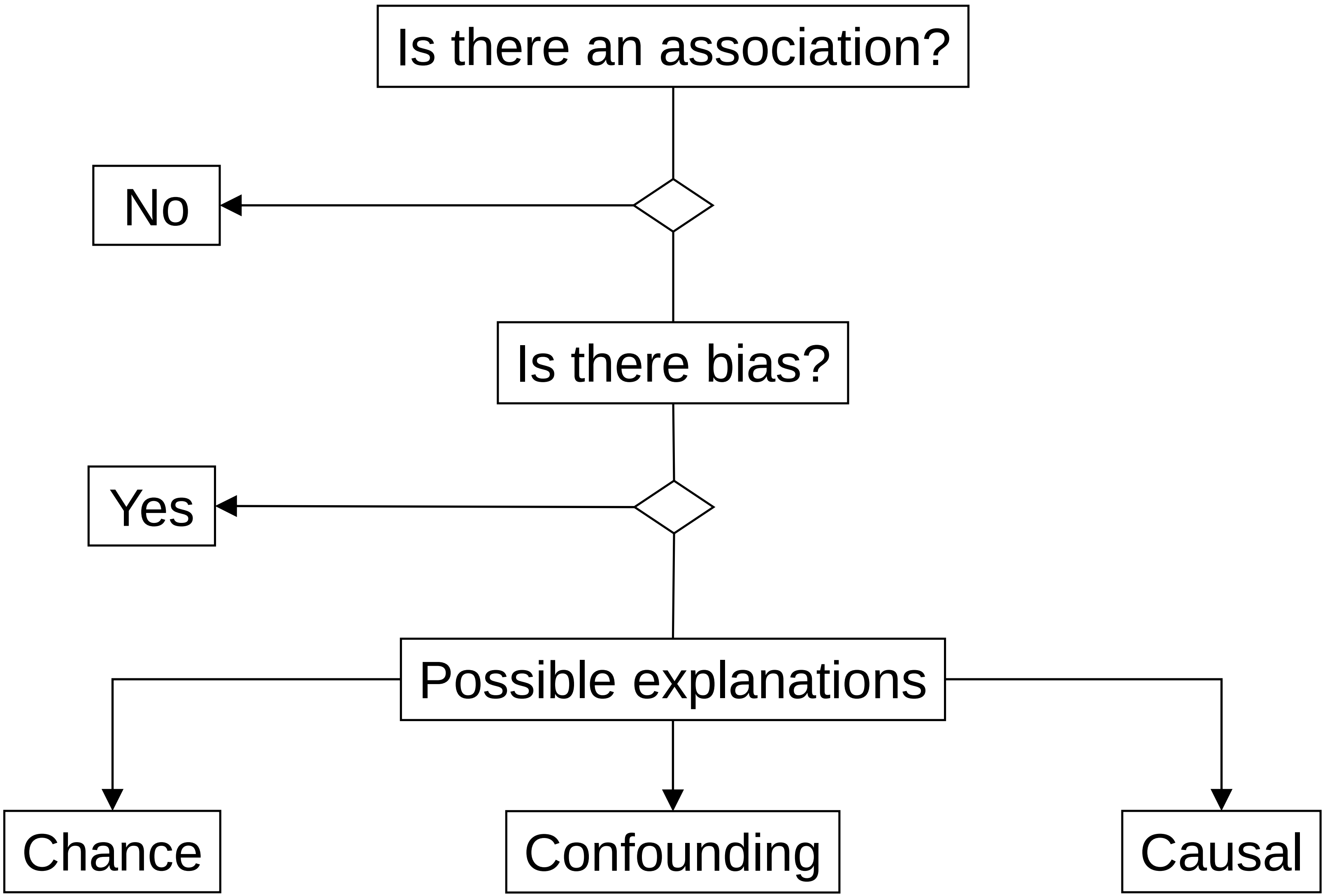
- Statistical methods (stratification, standardization, regression)

Why Is Confounding Important?

- Interferes with search for causal associations
- If association is not causal, intervention will not be effective

	Cross-sectional	Case-control	Cohort	Clinical trial
Selection bias:				
• *Nonresponse*	×	×	×	×
• *Loss to follow-up*			×	×
Information bias	×	×	×	×
Confounding	×	×	×	

Epidemiologic Reasoning



Criteria for Causality

Temporality*

- The cause must precede the effect in time

Strength of the association*

- Strong associations are more likely to be causal than weak associations

Dose-response effect*

- If higher levels of exposure result in higher risk of disease, the association is more likely to be causal

Consistency

- Repeated observation of the association in different populations under different circumstances supports causality

Biological plausibility

- Causality is supported if the association makes sense in the context of current biological knowledge

** Applied to findings of a single study*

Useful tools for study design and evaluation

- CONSORT (RCTs)
  - www.consort-statement.org/

- STROBE (observational studies)
  - https://www.strobe-statement.org/index.php?id=strobe-home