Introduction to Clinical Epidemiology

Mona Baumgarten, Ph.D.

July 15, 2002

Outline

- Measures of disease occurrence
- Measures of association
- Epidemiologic study designs
- Bias
- Confounding
- Causality

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

- In 1960, 651 men were examined as part of the Charleston Heart Study
- 49 men were found to have coronary heart disease
- Prevalence of CHD = ?

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
  - Have not been immunized against the disease

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

Example

- In 1960, 651 men were examined as part of the Charleston Heart Study
- 49 men were found to have coronary heart disease
- During the period 1960 to 1975, 116 men developed CHD
- Cumulative incidence of CHD = ?

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^{-1}
- 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, death, loss to follow-up, or the end of the study

Example

Subject	Years of Follow-up	Died
1	2	N
2	2	Y
3	1	N
4	1	N
5	1	Y
6	3	N
7	1	Y
8	1	Y
9	1	N
10	2	Y

Example

$$\text{Incidence rate} = \frac{5 \text{ deaths}}{15 \text{ person-years}}$$

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

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

Incidence and Prevalence

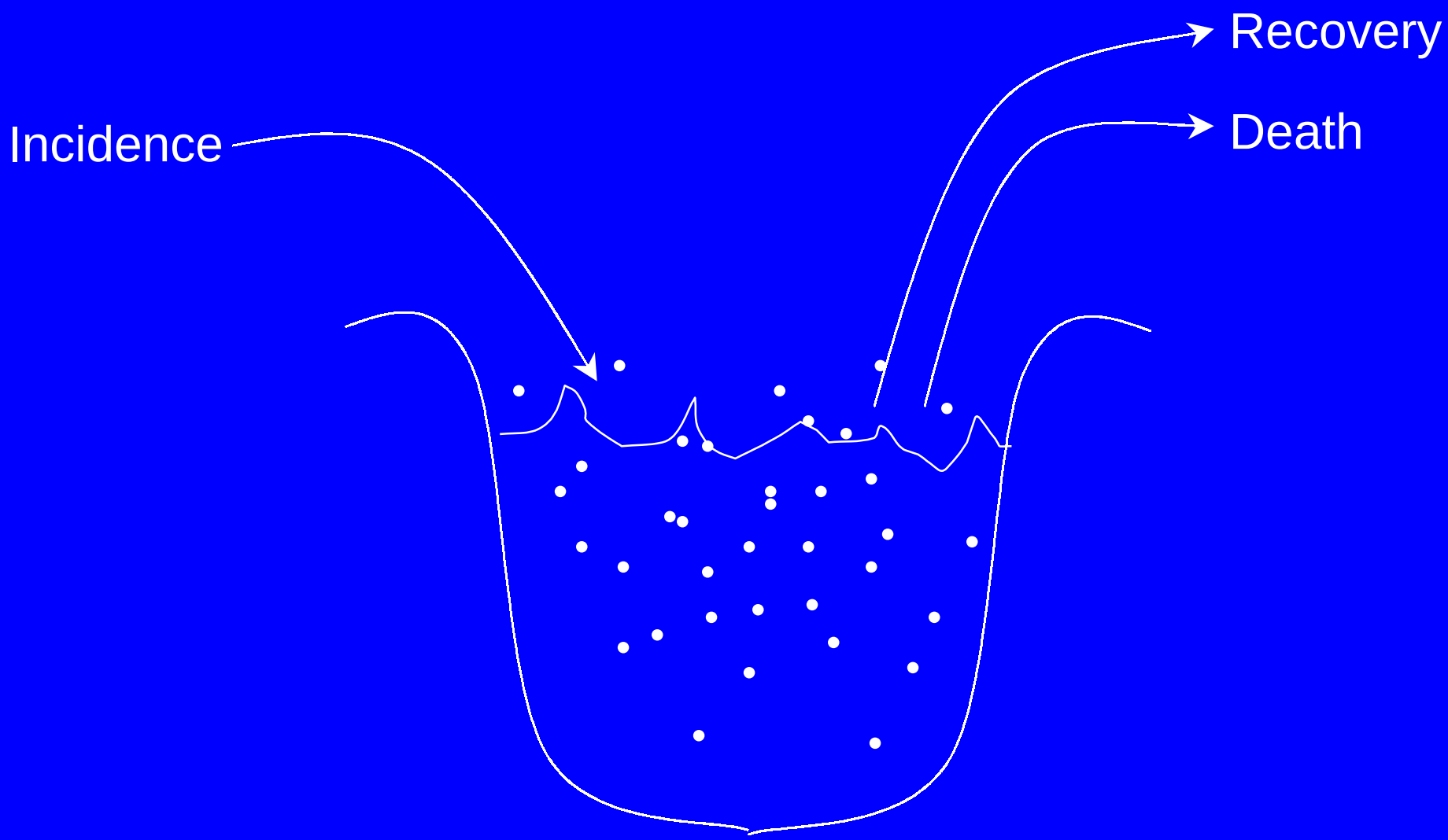
- Change in incidence reflects change in etiologic factors (risk factors or protective factors)
- Change in prevalence reflects change in incidence or duration or both

Incidence and Prevalence

Prevalence $\approx$
incidence rate $\times$ average duration of disease

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

Prevalence



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

Smoking \ CHD	Yes	No	
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

Annual Incidence of CHD Among Women Aged 50 to 69

Smokers	10%
Nonsmokers	2%

Relative Risk = ?

Odds

- Ratio of the probability that an event will occur to the probability that the event will not occur

Example

- 20 smokers develop chronic bronchitis while 30 do not
- Odds of bronchitis = ?
- Probability of bronchitis = ?

Odds Ratio

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

Odds Ratio

Smoking	CHD: Yes	CHD: No
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 risk of disease
- When OR < 1, the exposure is a protective factor for the disease, i.e., decreases the risk of disease

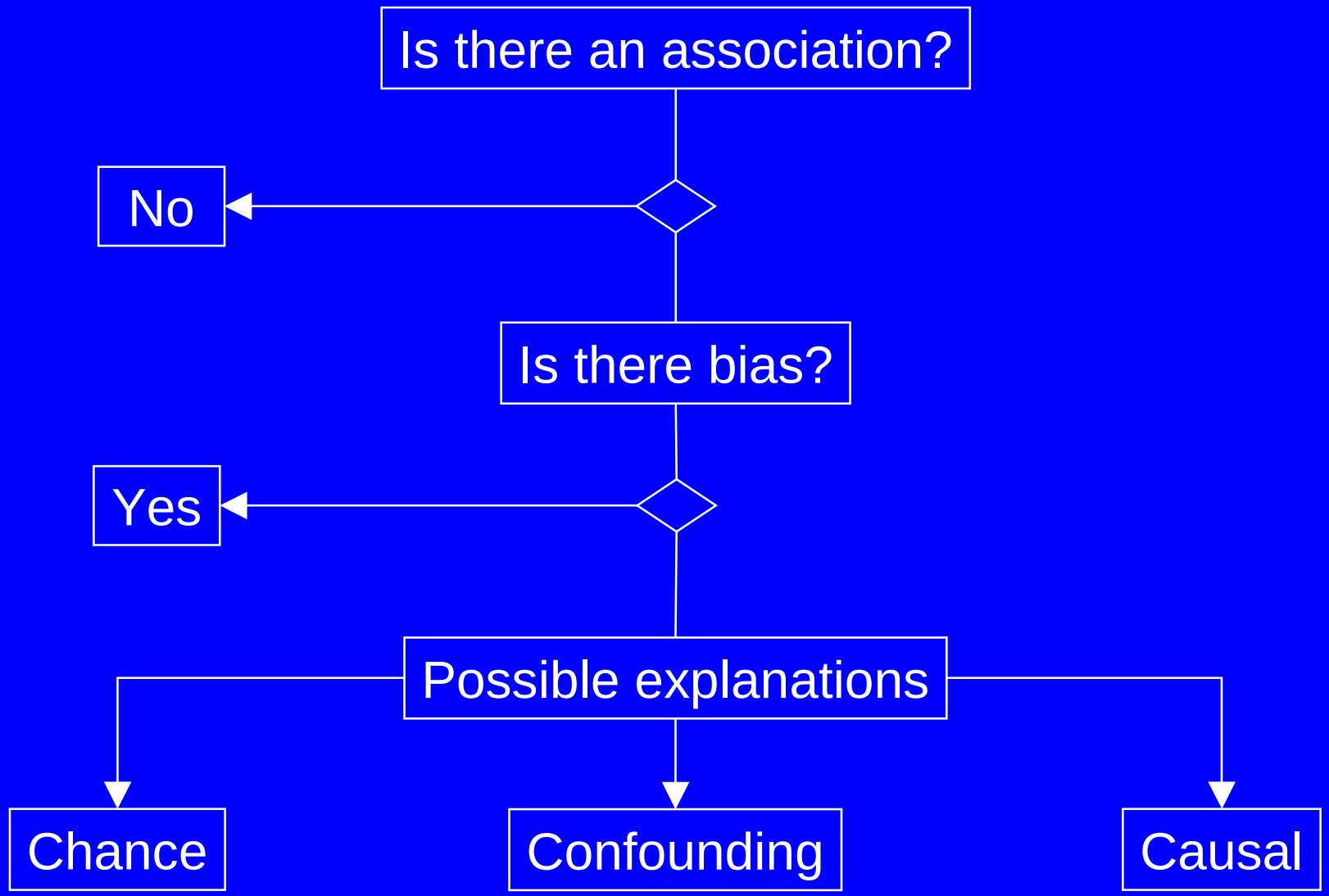
Odds Ratio

- 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) *and*
  - the control group is representative of the general population with respect to the frequency of exposure

Example

- In a case-control study, 25 of 100 CHD cases were smokers while 10 of 100 controls were smokers
- Odds ratio = ?

Epidemiologic Reasoning



Epidemiologic Study Designs

Experimental Studies

- Clinical Trials
- Other Experimental Studies

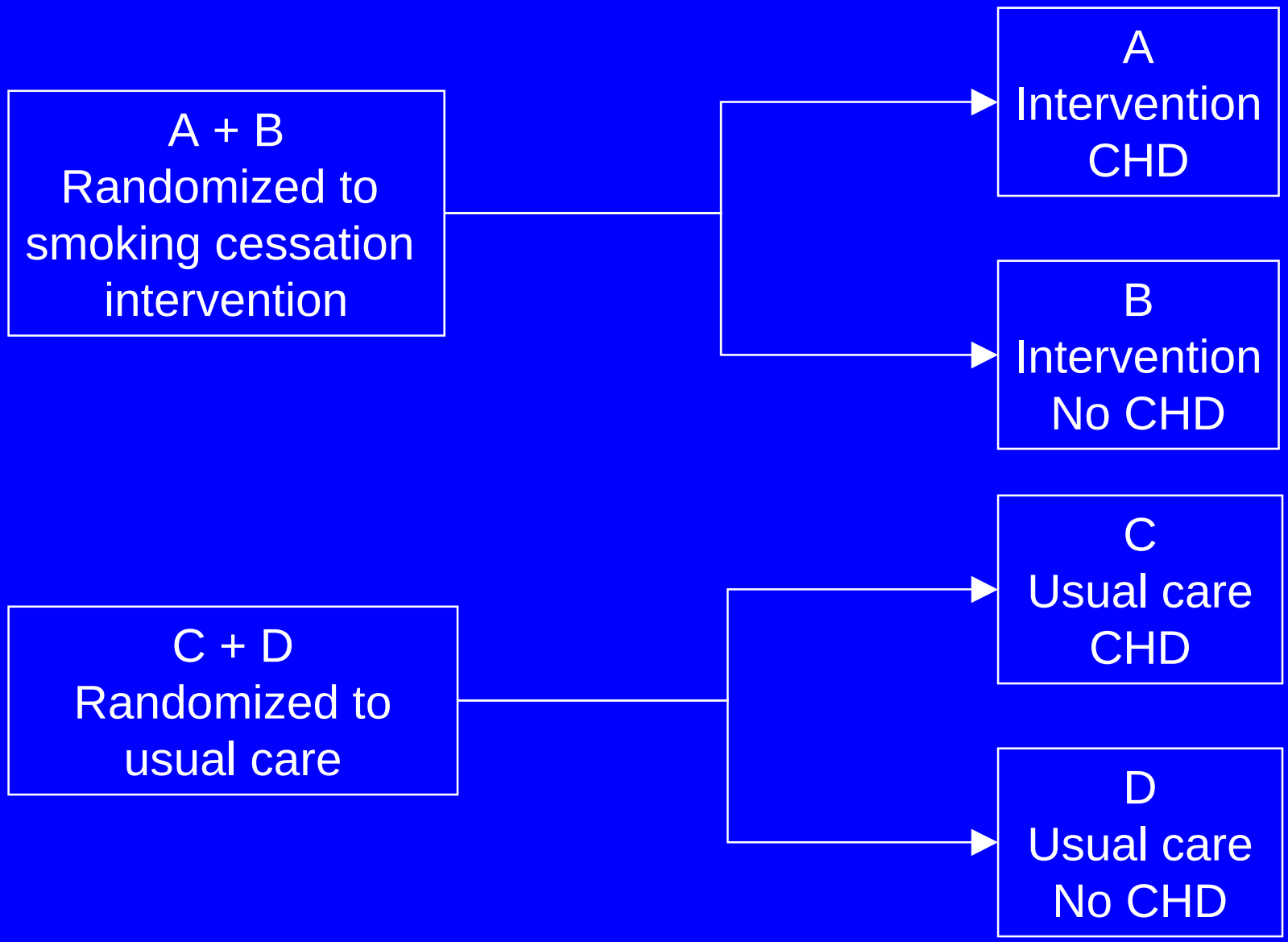
Observational Studies

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

Clinical Trials

- Treated and untreated subjects are followed over time to determine whether they experience the outcome
- Assignment to treatment or non-treatment is by randomization

Clinical Trials



Clinical Trials

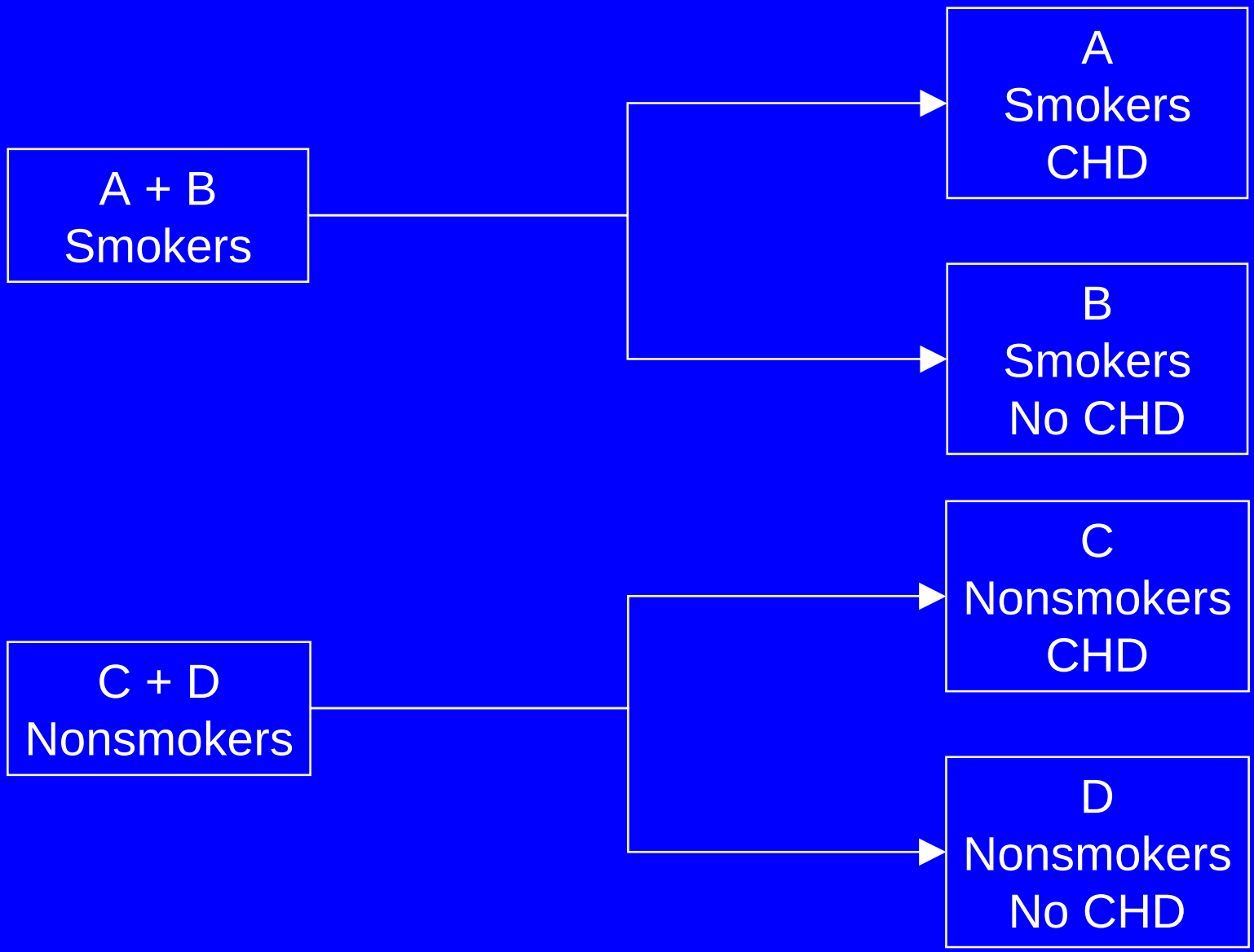
Intervention	CHD: Yes	CHD: No	
Yes	a	b	a+b
No	c	d	c+d
	a+c	b+d	

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

Cohort Studies

- Exposed and unexposed subjects without disease are followed over time to determine whether they experience the outcome
- Experimental studies (randomized trials) are a special case of the cohort study

Cohort Studies



Cohort Studies

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

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

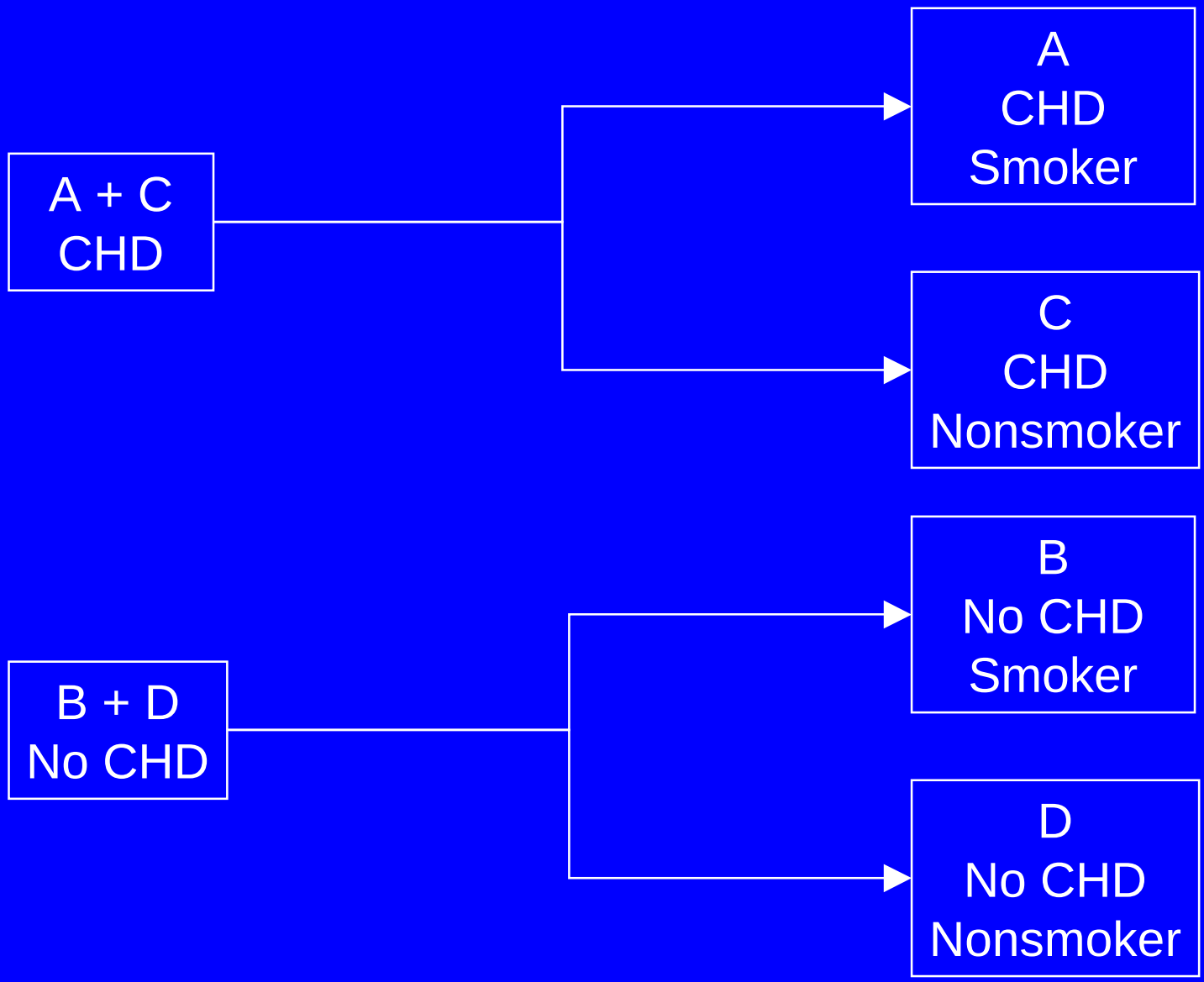
Case-Control Studies

- Compare exposure among persons with the disease (cases) to exposure among persons without the disease (controls)
- Most commonly used epidemiologic study design despite numerous potential biases

Case-Control Studies

- More efficient than the equivalent cohort study
- Makes it possible to study rare diseases

Case-Control Studies



Case-Control Studies

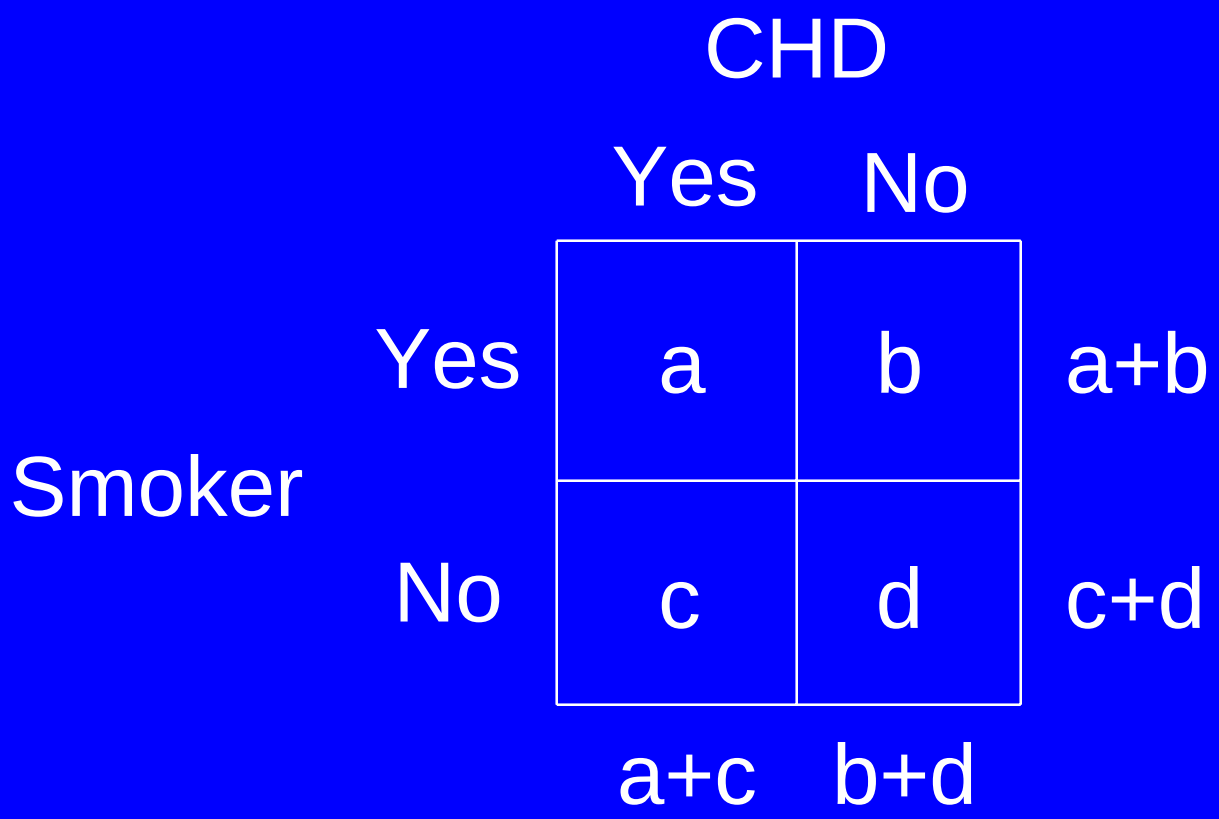
Smoking	CHD: Yes	CHD: No
Yes	a	b
No	c	d
	a+c	b+d

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

Cross-Sectional Studies

- Survey of a sample of the population in which the status of individuals with respect to one or more characteristics is assessed at one point in time

Cross-Sectional Studies



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

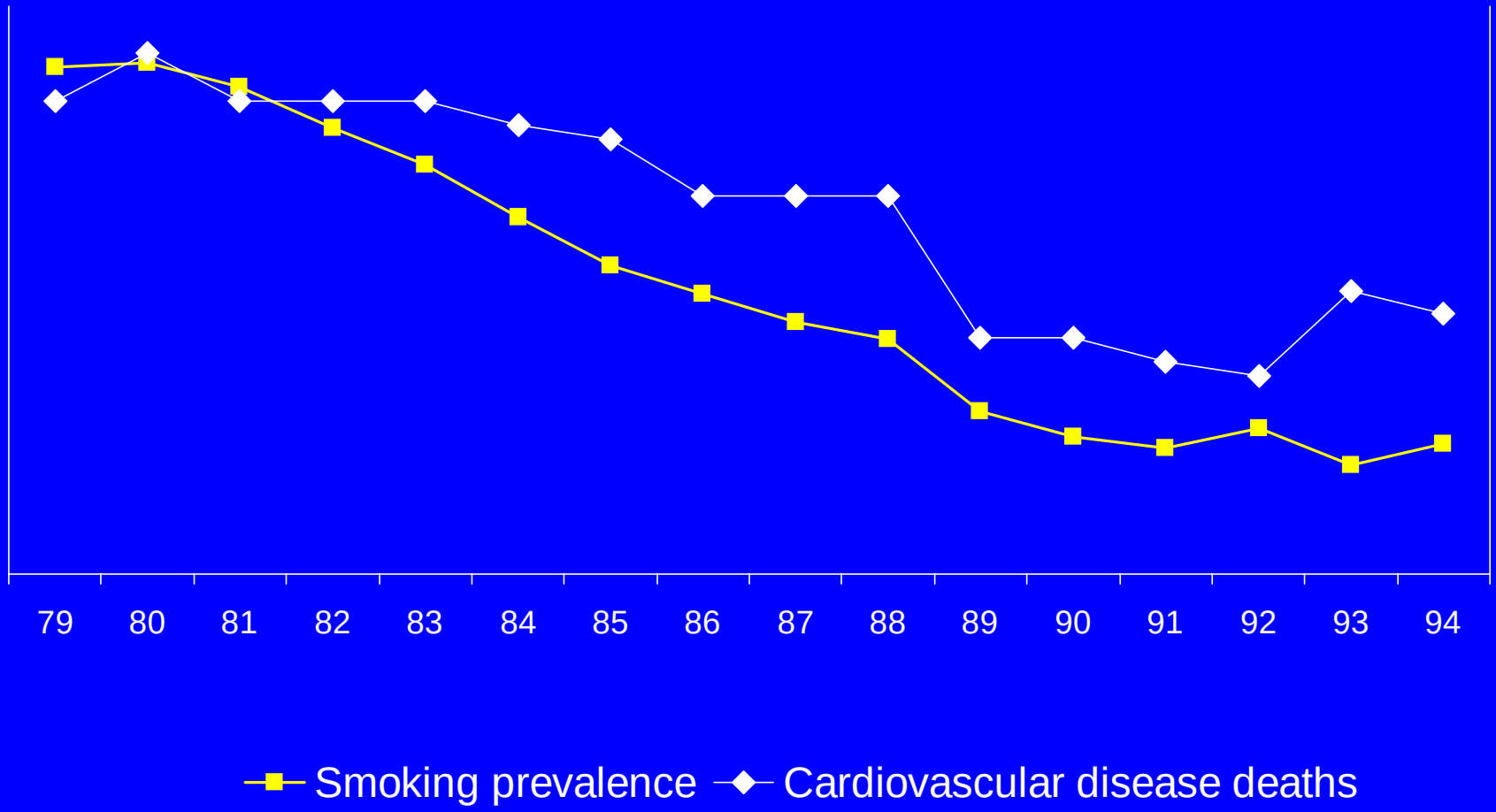
Ecologic Studies

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

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

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



Case Series

- Studies without a comparison group
- All study subjects have the disease (or the exposure)
- Impossible to make inferences about causality

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

	Cross-sectional	Case-control	Cohort	Clinical trial
Selection bias:				
• *Sampling frame*	✓			
• *Nonresponse*	✓	✓	✓	
• *Self-selection*	✓	✓	✓	✓
• *Detection*		✓		
• *Loss to follow-up*			✓	✓
Information bias	✓	✓	✓	✓
Confounding	✓	✓	✓	

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 choice of sampling frame
- Bias related to nonresponse
- Bias related to self-selection
- Detection bias
- Bias related to loss to follow-up

Choice of Sampling Frame

Example:

- In a prevalence survey, the telephone directory was used to select study subjects

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

Self-Selection

- Reasons for self-selection may be related to important study variables
- A particular problem in studies carried out among volunteers

Self-Selection

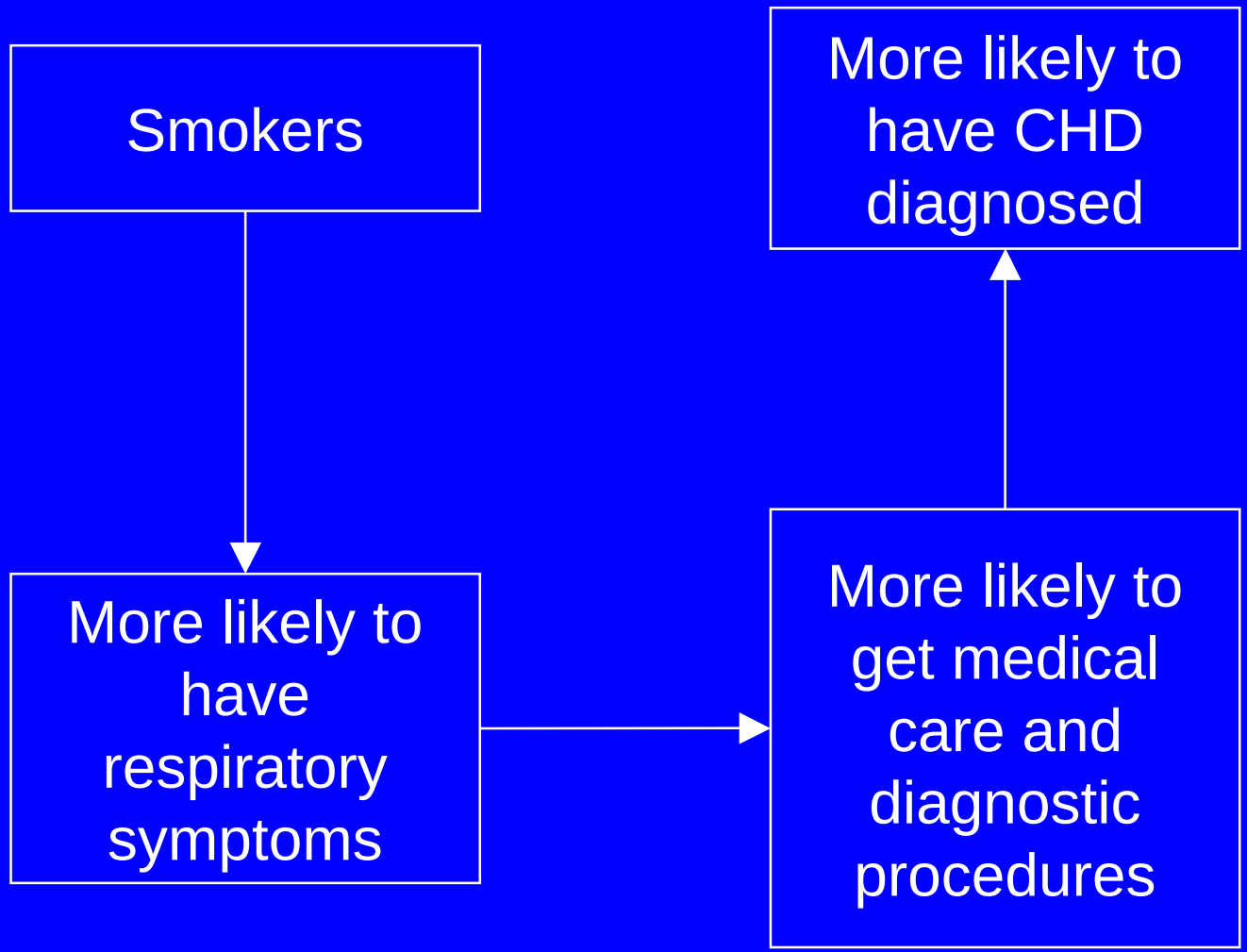
Example:

- Smokers who volunteer to be in a study to evaluate the effectiveness of smoking cessation may be more motivated or more desperate than other smokers

Detection Bias

- May arise in case-control studies
- Occurs if people with the exposure are more likely to receive medical surveillance and, therefore, to have the disease identified

Detection Bias



Loss to Follow-Up

- In cohort studies and clinical 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

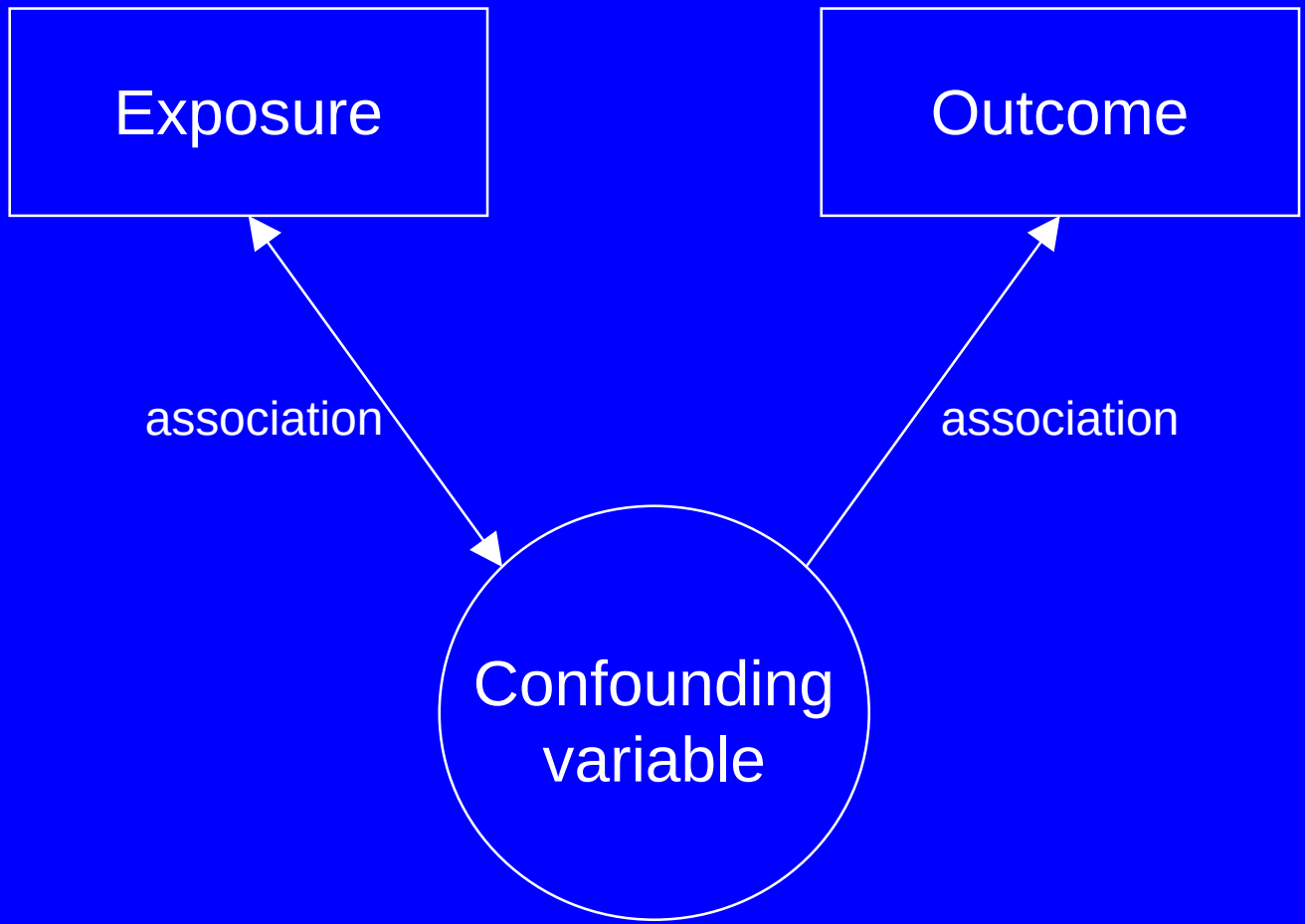
Confounding

- 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

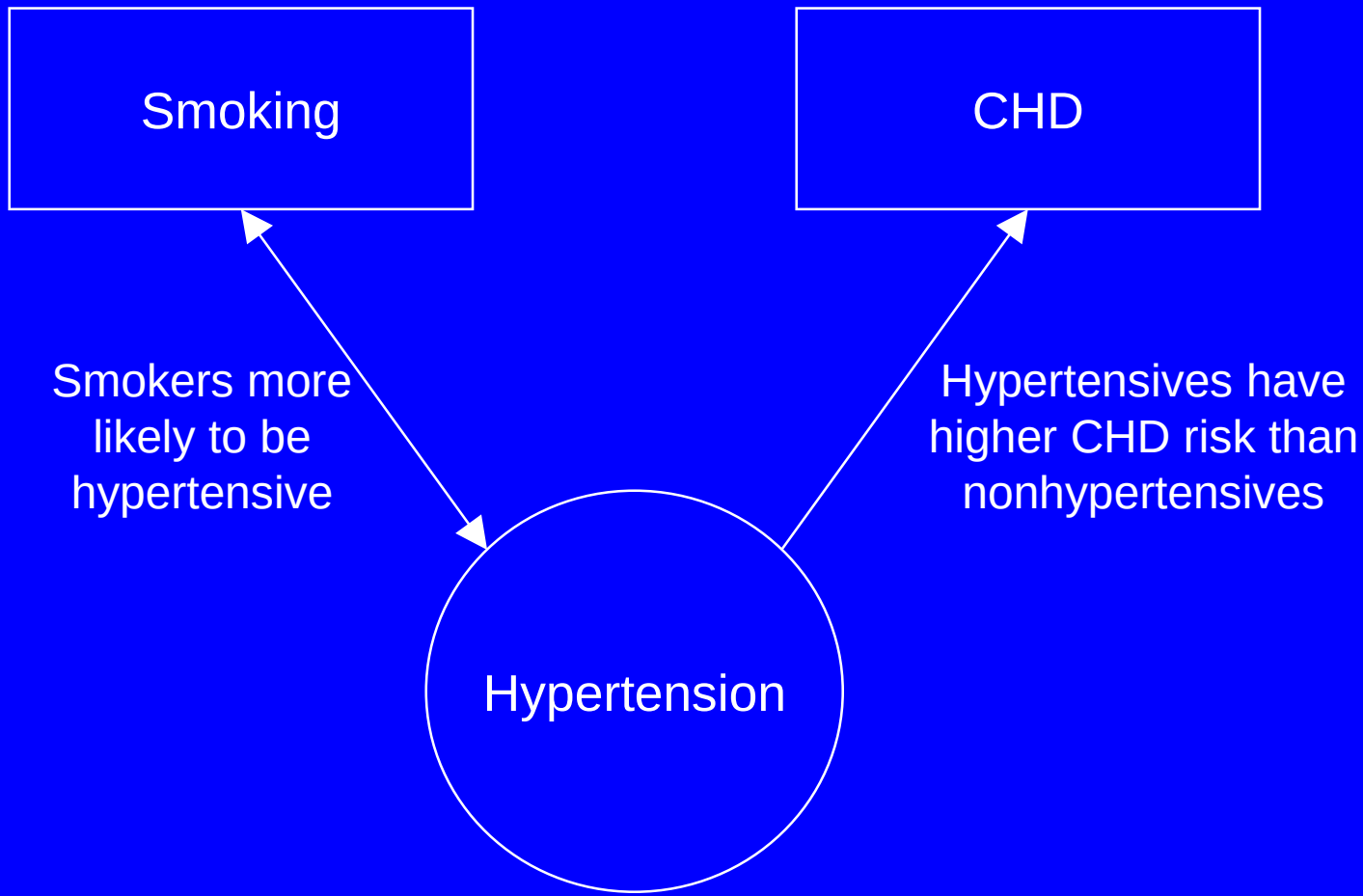
Confounding

- 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



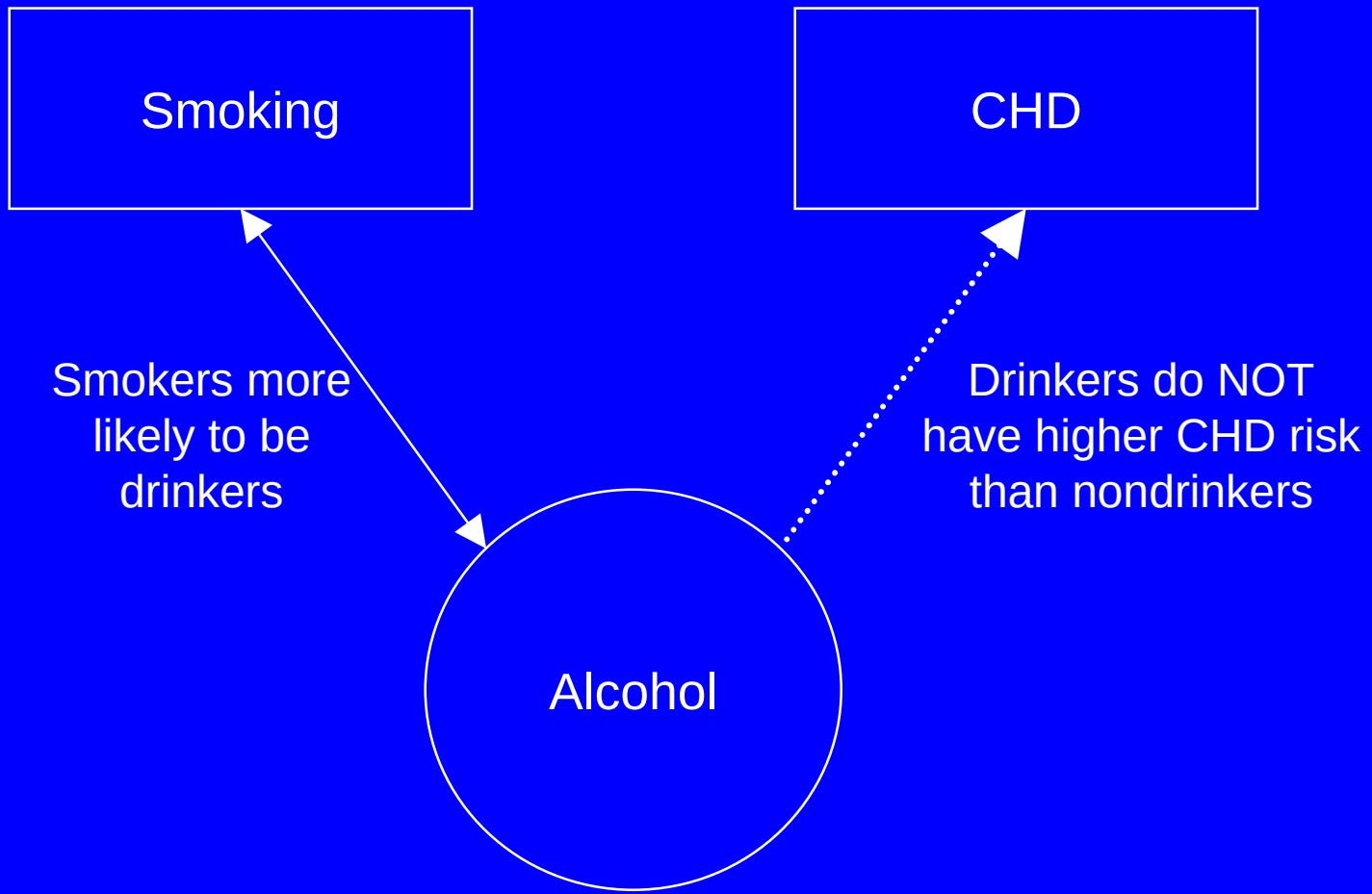
Example


Smoking
CHD
Smokers more likely to be hypertensive
Hypertensives have higher CHD risk than nonhypertensives
Hypertension

Example

- Suppose you wish to study the effect of smoking on the risk of CHD
- Hypertension is distributed differently among smokers and nonsmokers
- Hypertension is a risk factor for CHD
- Therefore, hypertension is a potential confounder

Example


Smoking
CHD
Smokers more likely to be drinkers
Drinkers do NOT have higher CHD risk than nondrinkers
Alcohol

Example

- Suppose you wish to study the effect of smoking on the risk of CHD
- Alcohol drinking is distributed differently among smokers and nonsmokers
- Alcohol is not a risk factor for CHD
- Therefore, alcohol is not a potential 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)

When Not to Adjust

- When the variable lies in the causal chain between the exposure and the outcome

When Not to Adjust

- When the association between the exposure and the outcome differs at different levels of the variable

- *Example:*

  RR smoking/CHD in men: 6.2

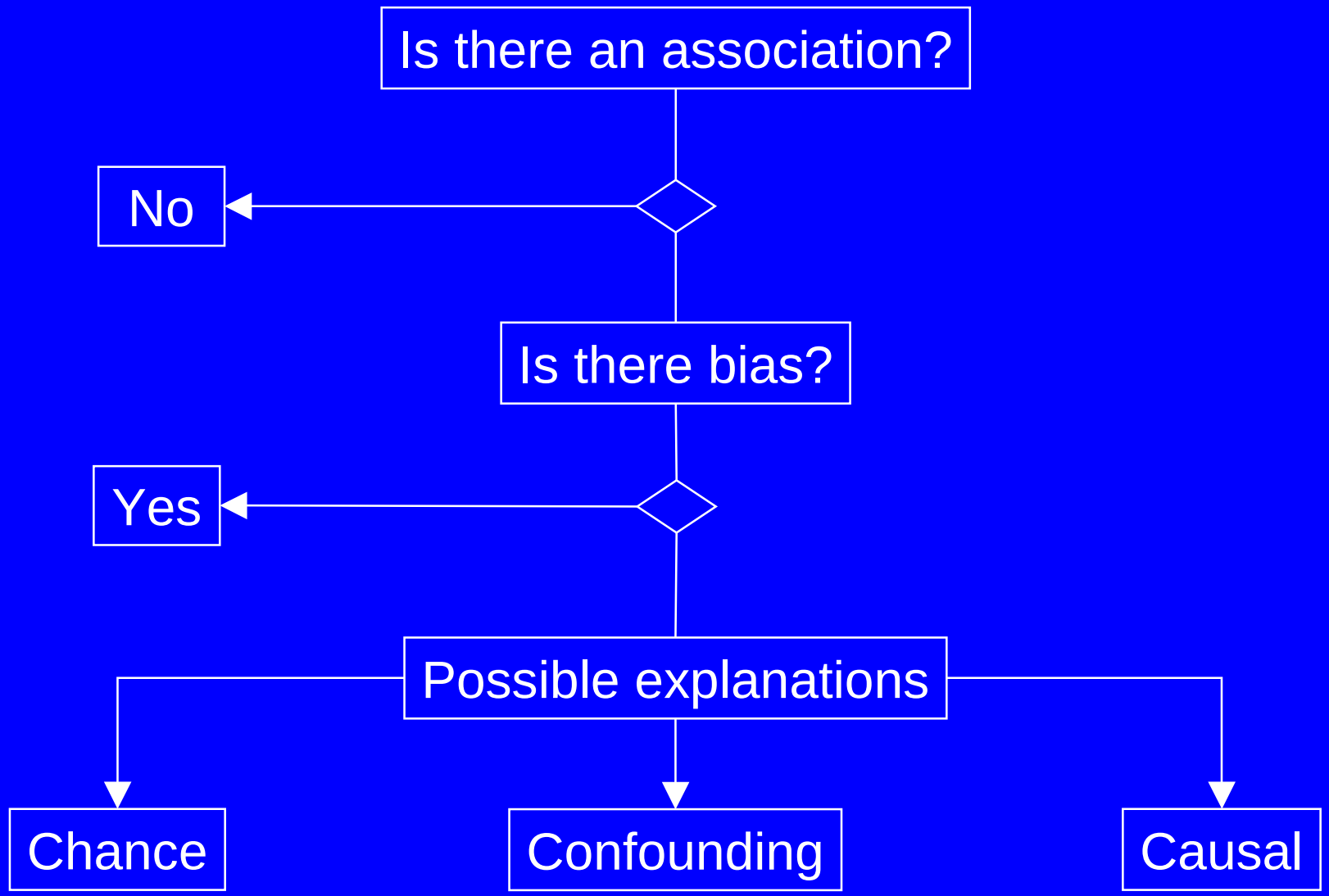
  RR smoking/CHD in women: 2.1

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:				
• *Sampling frame*	✓			
• *Nonresponse*	✓	✓	✓	
• *Self-selection*	✓	✓	✓	✓
• *Detection*		✓		
• *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*