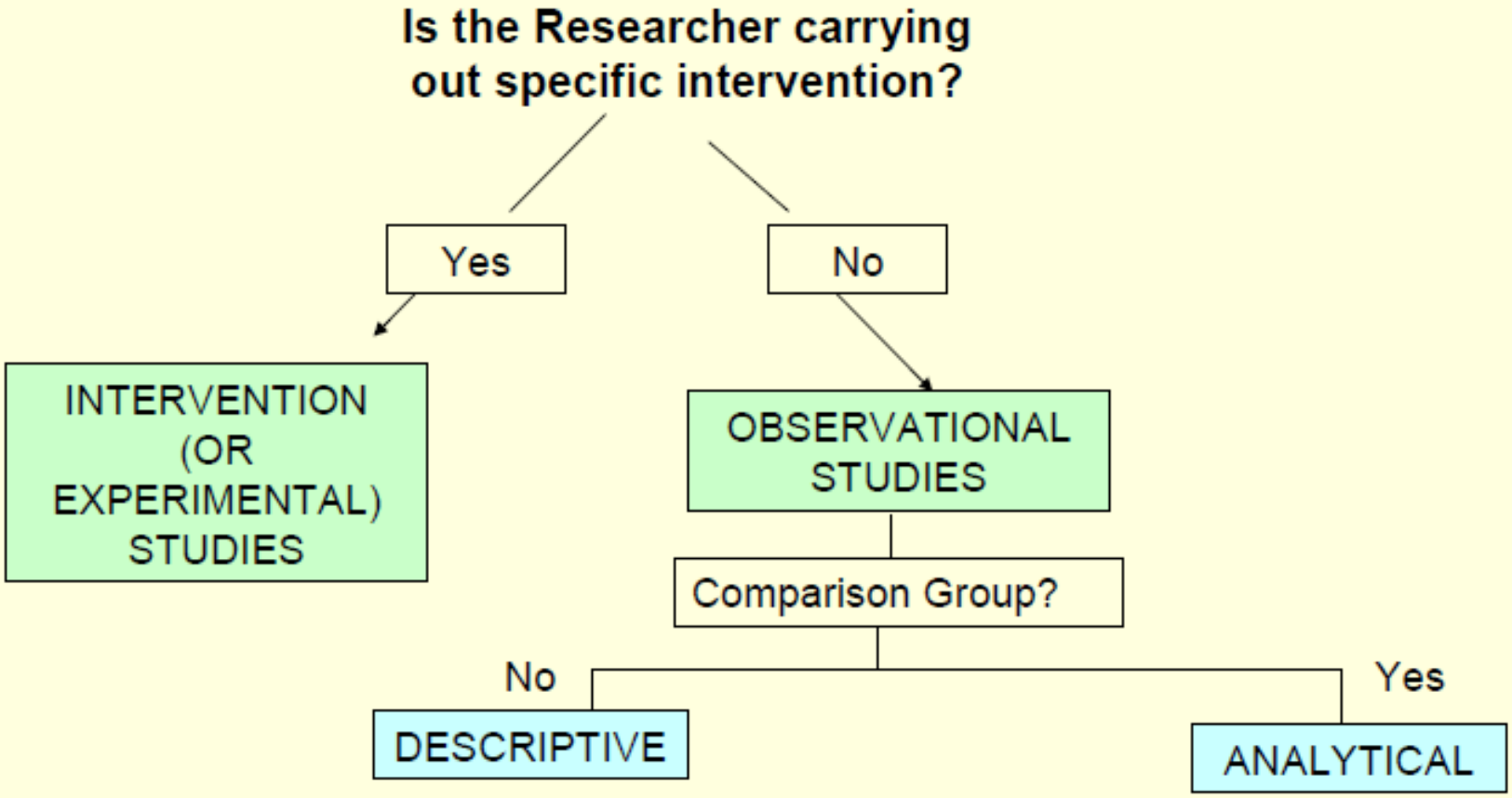
Design Strategies and Measurement of association

Dr Olarewaju Sunday

What are the objectives of epidemiology?

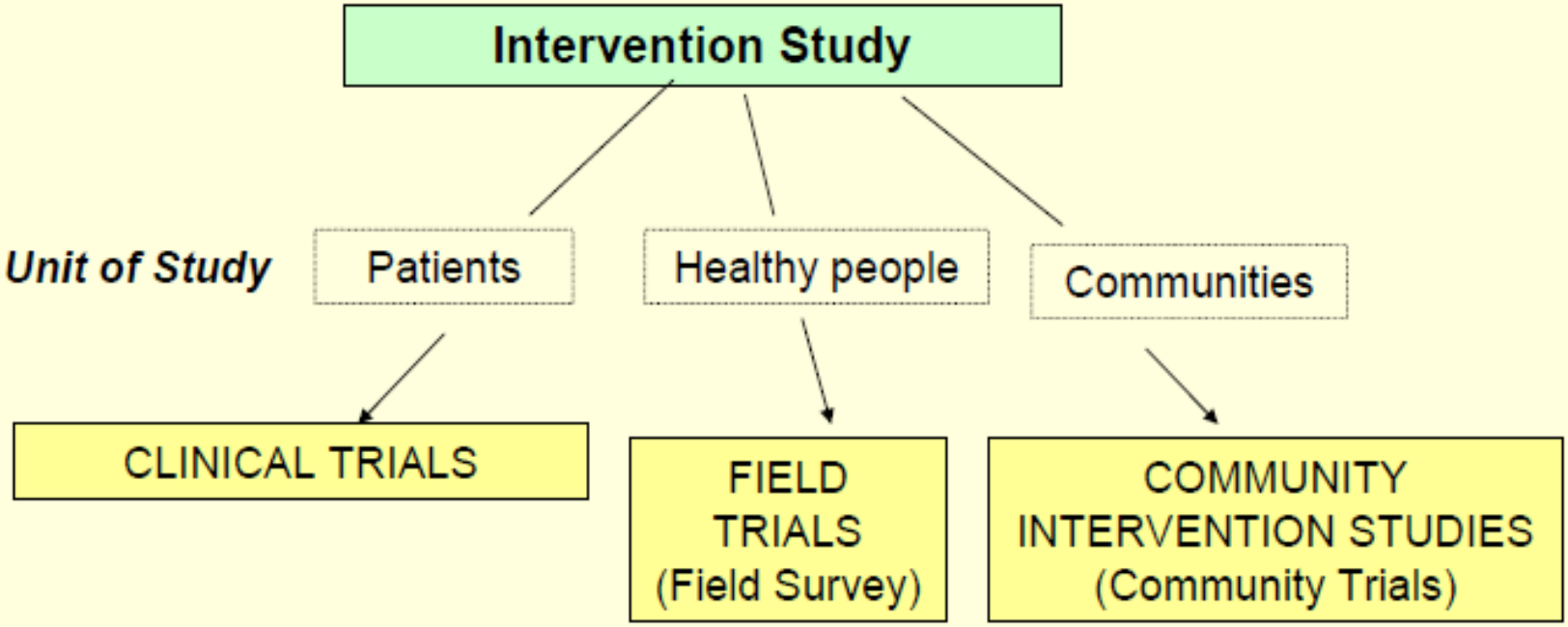
1. Quantify burden (Time, Place & Person)- Measurement tools
2. Identify the determinant factors
3. Establish causal relationship between exposure and outcome- **Study designs**

Epidemiologic Study Designs



Experimental studies

Study Designs: Intervention Studies



Experiment may be carried out in one of the following settings

1) **Clinical setting**- carried out on patients, for testing a **new treatment**, mostly 'Randomized CLINICAL trial'.

2) **Field Trial** – carried out in the community- on healthy individuals- mostly for testing **prophylactic agents like a vaccine**

-This can employ a randomized controlled design or a non randomized design

3) **Community Trial** –carried out in a community-the **intervention** has to be made at **public level**

Basic Steps in Randomized Controlled Trial

- The basic steps in conducting a RCT include the following:
- 1. Drawing up a protocol
- 2. Selecting reference and experimental populations
- 3. Randomization
- 4. Manipulation or intervention
- 5. Follow-up
- 6. Assessment of outcome

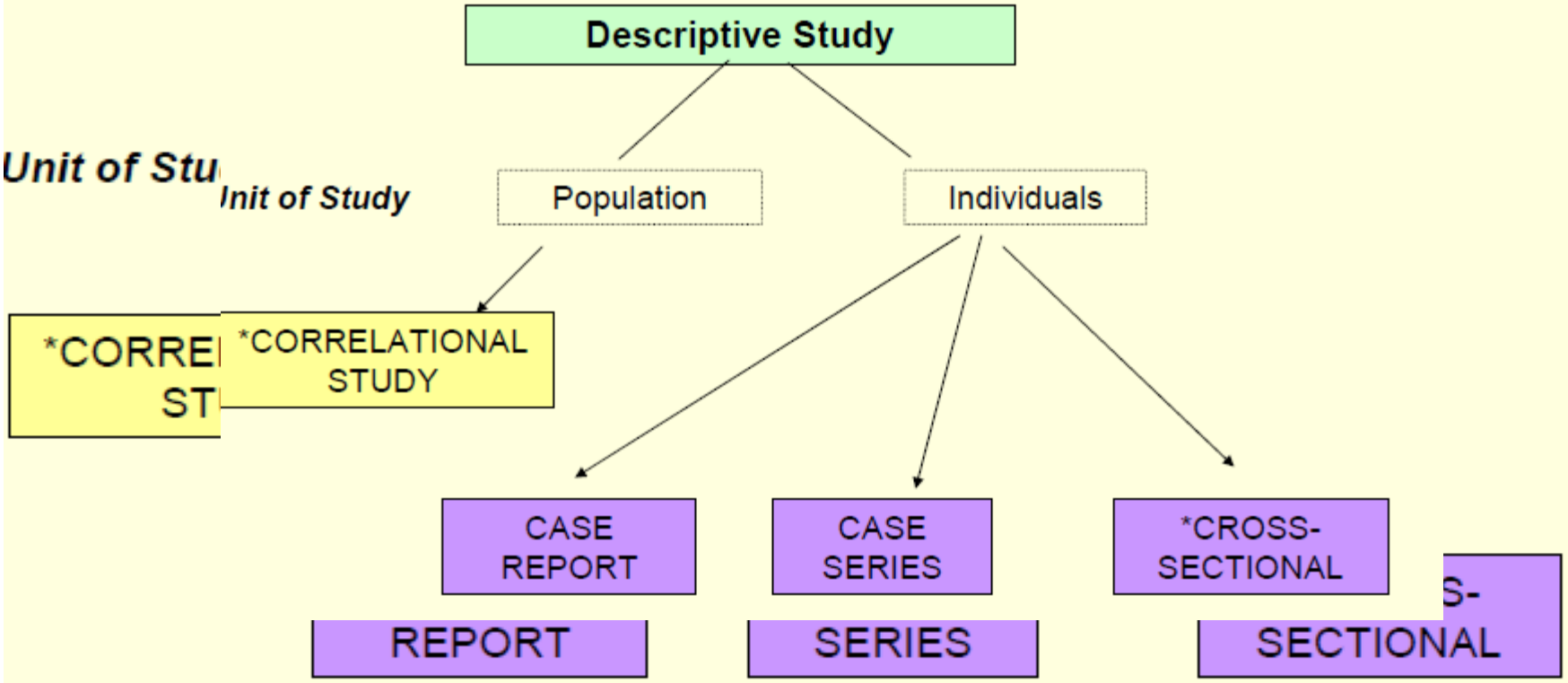
Continue in next session.

Blinding/Confounders

Observational studies

Study Designs: Descriptive Studies

Study Designs: Descriptive Studies



Case Reports & Case Series

- An important interface between clinical medicine & epidemiology
- Useful in generating hypothesis
- Cannot be used to test for the presence of valid association because they lack appropriate comparison group

Correlational Studies

- Approach: Measures that represent characteristics of entire populations are used to describe the disease in relation to some factor of interest (such as age, calendar time, food consumption, drug use, and utilisation of health services)
- Usefulness: Generate hypothesis about possible exposure-outcome relationship

Ecologic Correlation of Breast Cancer Mortality and Dietary Fat Intake

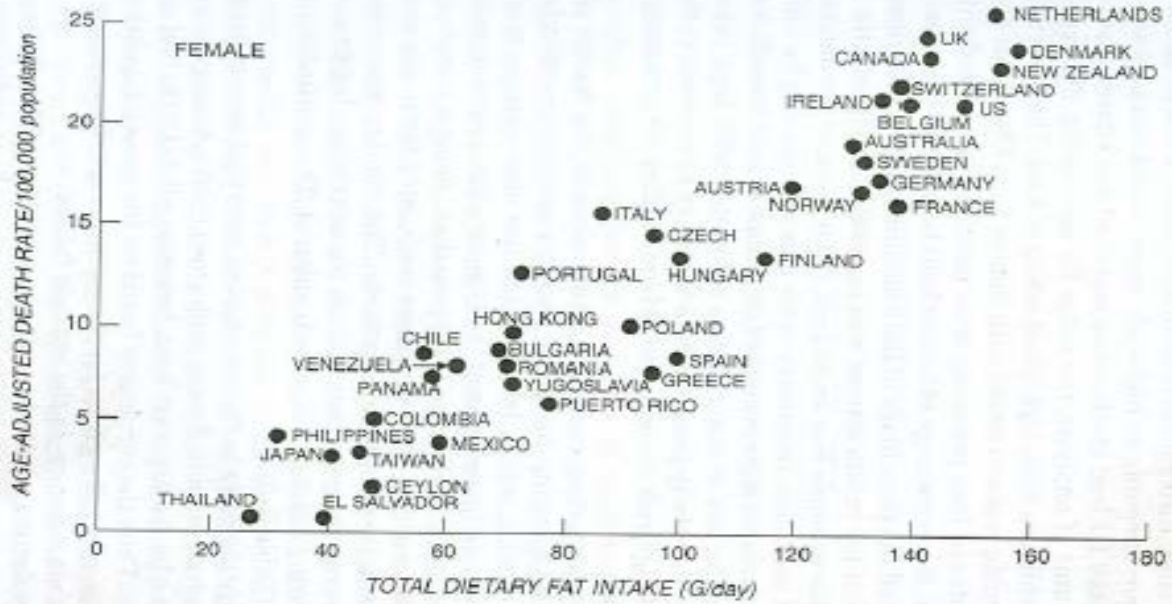


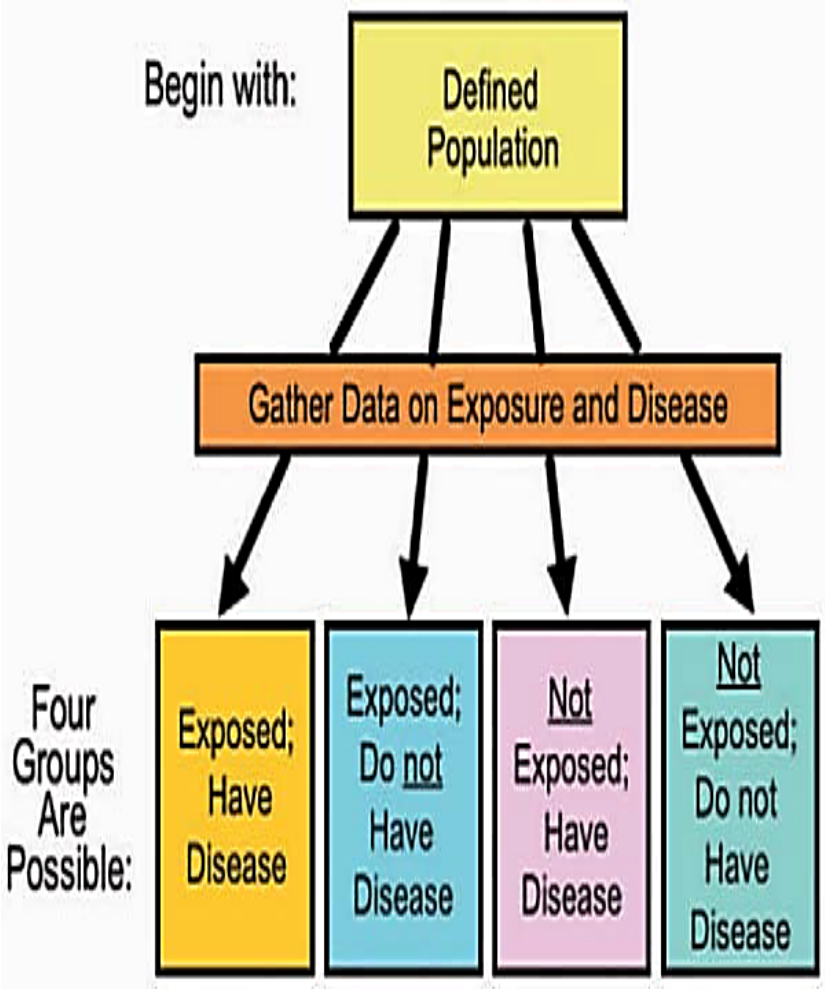
Figure 6–3 Ecologic correlation of breast cancer mortality and dietary fat intake. *Source:* Reprinted from Carroll, K.K., Experimental Evidence of Dietary Factors and Hormone-Dependent Cancers, *Cancer Research*, Vol. 35, p. 3379, with permission of Waverly Press, Inc., © 1975.

Corelational Studies: Strength & Weaknesses

- Strength: Quick, easy and inexpensive to conduct – using existing (secondary) data
- Weakness:
  - Inability to link exposure with outcome at individual level (problem of "ecological fallacy")
  - Inability to control for the effects of potential confounding factors

Cross sectional descriptive studies

Cross sectional studies



- We define a population and determine the presence or absence of exposure and the presence or absence of disease for each individual
- Each subject then can be categorized into one of four possible subgroups

Design of a cross-sectional study

Cross sectional studies

The findings can be viewed in a 2 × 2 table, as seen in Figures which also show the two approaches to interpreting the findings from such studies.

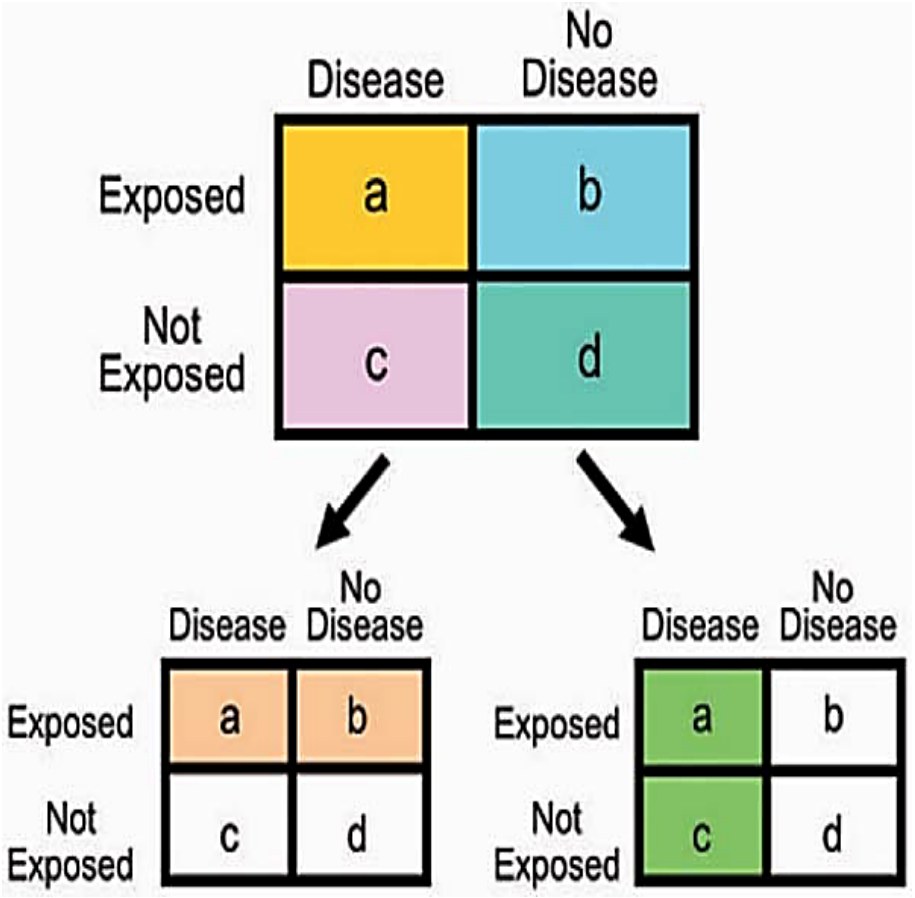


FIGURE 2-Design of a cross-sectional study: II

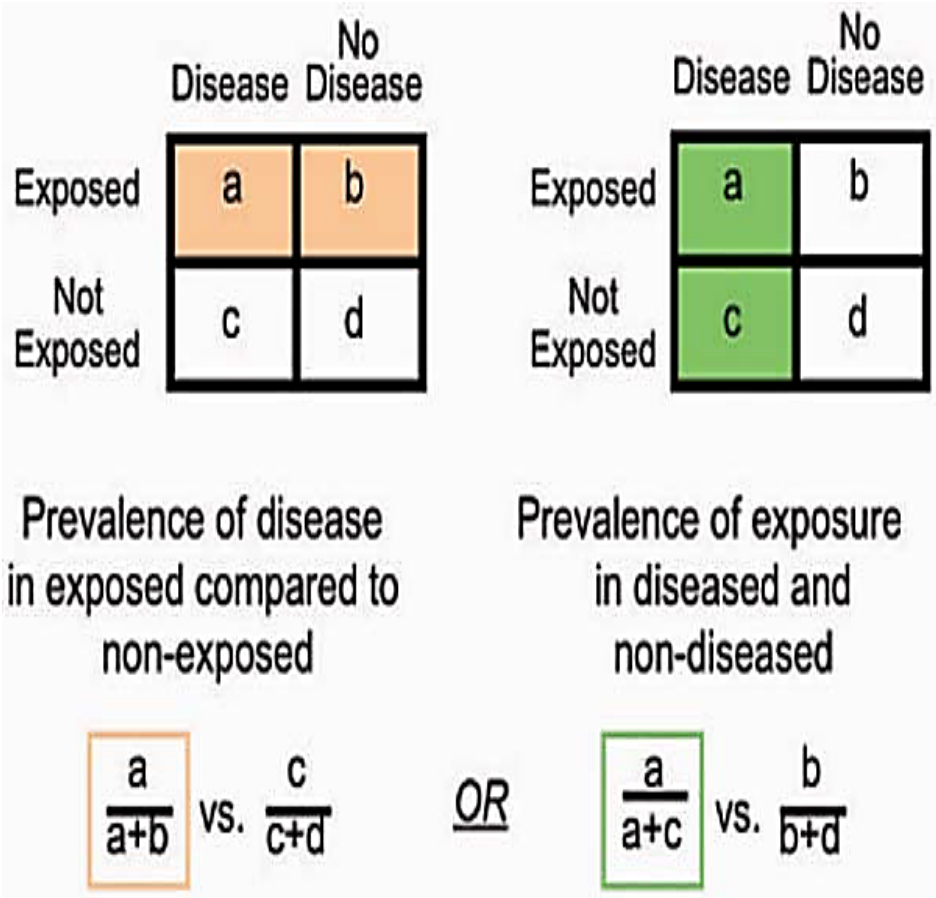


FIGURE 3-Design of a cross-sectional study: III.

Cross-Sectional Study

- Approach: Exposure and outcome status are assessed simultaneously among individuals
- Useful in determining prevalence of health problem and generating hypothesis
- When current values of exposure variables are unalterable over time (e.g. blood group or eye colour), Cross-sectional survey can be used to test hypotheses (analytic study)

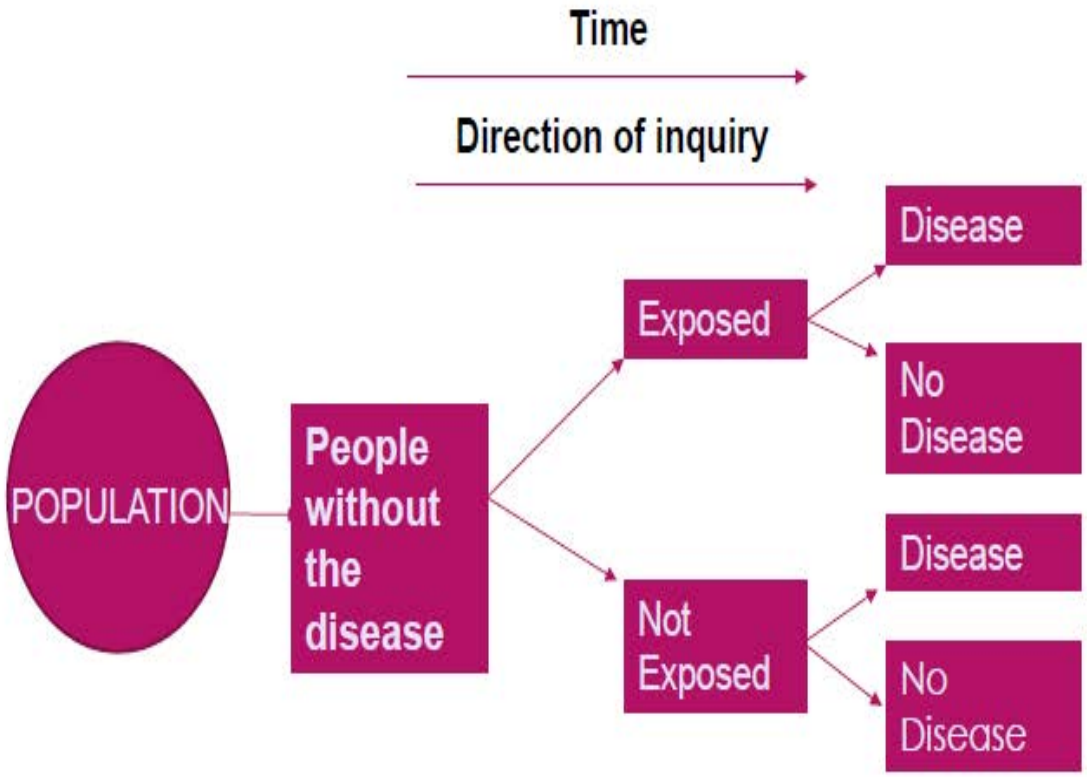
Cross-Sectional Study: Weaknesses

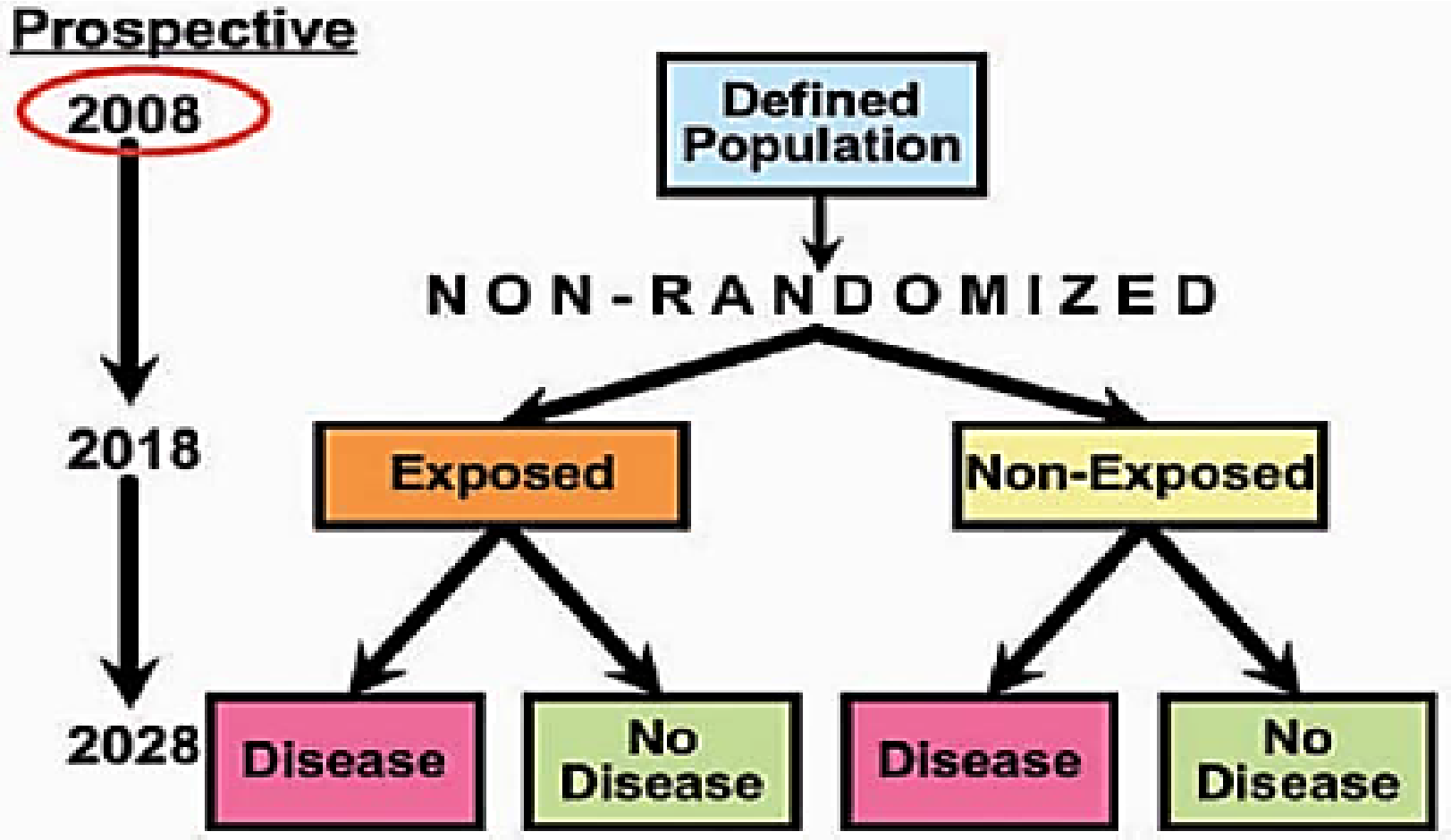
- Usefulness:
  - Assessment of health status & health care needs
  - Provide information on prevalence of disease or other health outcomes
- Weaknesses:
  - Prevalent cases (rather than incident are used) – difficult to sort out factors associated with risk of disease from factors associated with survival
  - Challenge of studying diseases of low frequency
  - Difficult to determine whether the exposure preceded or resulted in the disease (as exposure & outcome status are assessed at a single point in time)

Cohort study

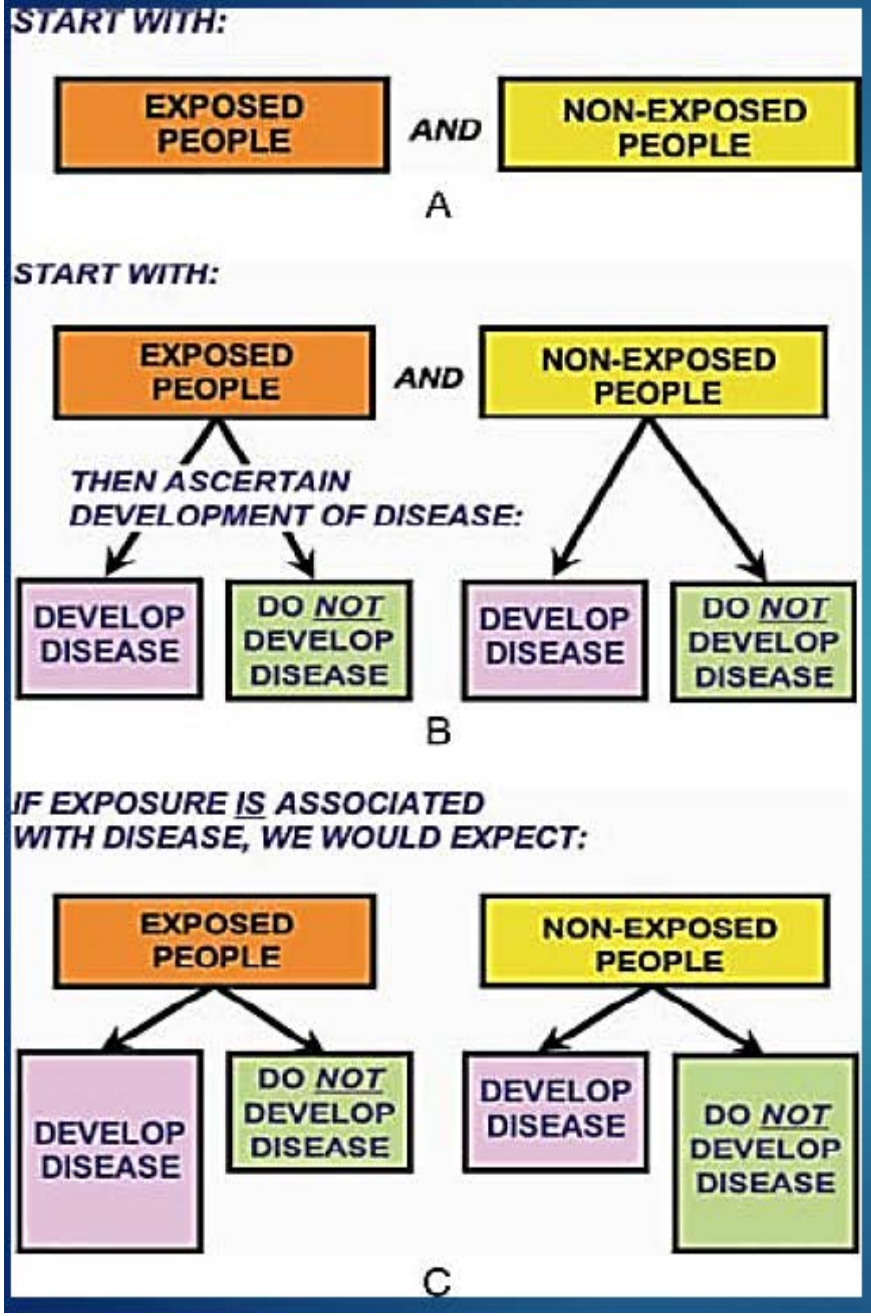
Analytical study

Design of Cohort Study



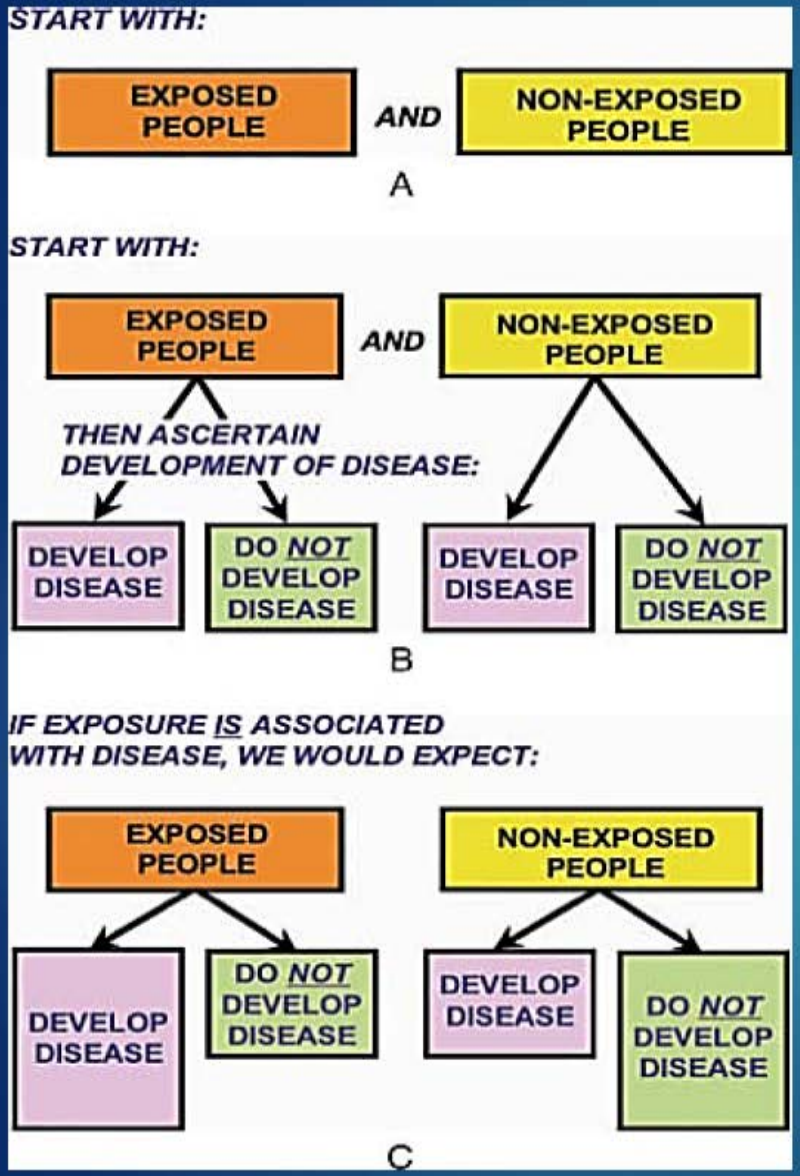


Time frame for a hypothetical prospective cohort study begun in 2008



Cohort Study Design

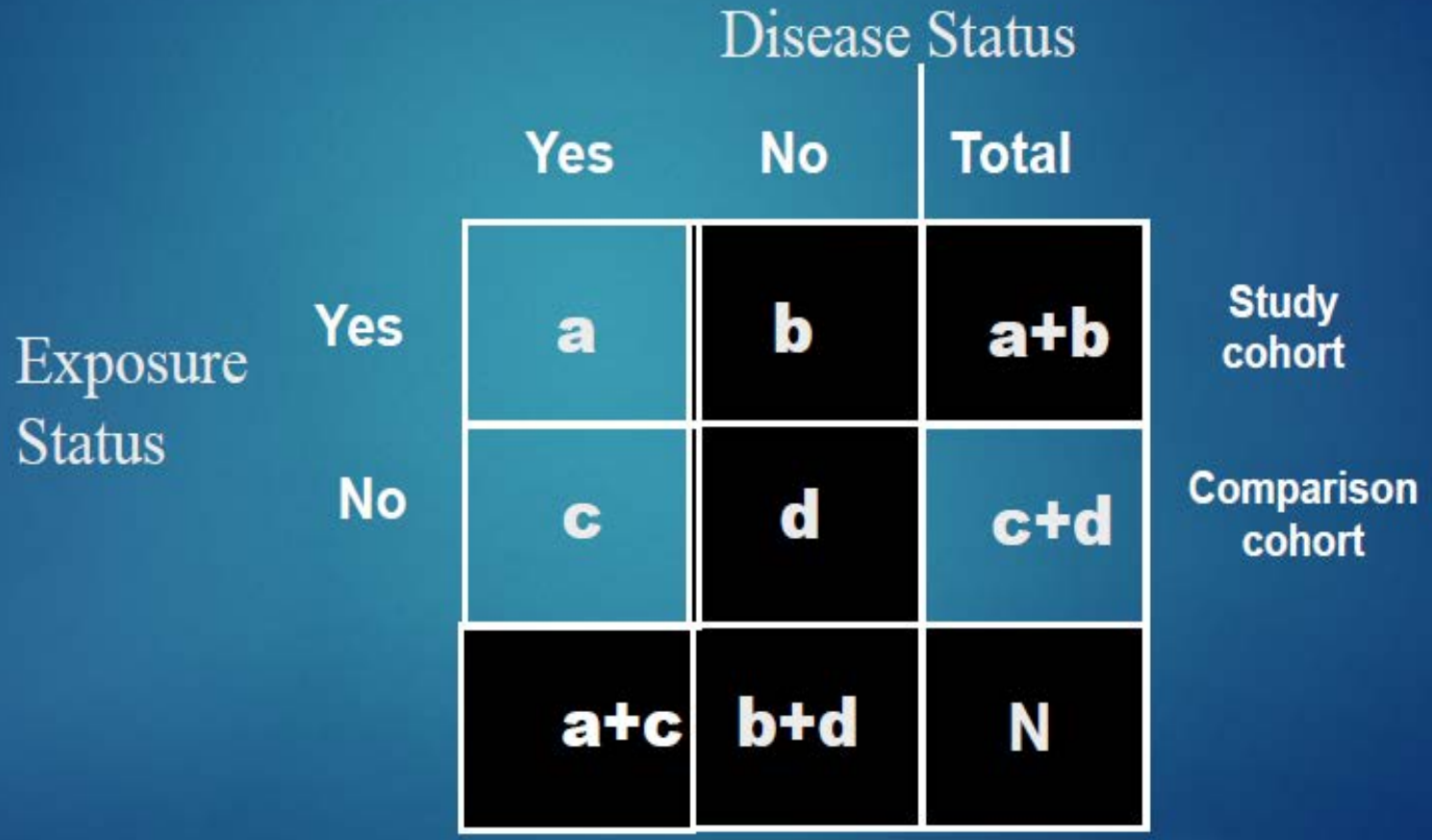
- Design of a cohort study.
- **A**, Starting with exposed and nonexposed groups.
- **B**, Measuring the development of disease in both groups.
- **C**, Expected findings if the exposure is associated with disease



Cohort Study Design

- Design of a cohort study.
- **A**, Starting with exposed and nonexposed groups.
- **B**, Measuring the development of disease in both groups.
- **C**, Expected findings if the exposure is associated with disease

Incidence rates of outcome



Relative risk

- Quantifies magnitude of the association between exposure and disease
- Varies from 0 to infinity
- **RR=1: no association**
- **RR>1: exposure is a risk factor for disease; Increase \risk for disease**
- **RR<1: exposure decreases the risk for disease**

Example:

- RR=2.0 can be interpreted as two fold increase in risk
- RR=0.7 can be interpreted as 30% decrease in risk

Example

Smoking	Lung cancer		Total
	YES	NO	
YES	70	6930	7000
NO	3	2997	3000
	73	9927	10000

Find out RR and AR for above data

- Incidence of lung cancer among smokers

  70/7000 = 10 per 1000

- Incidence of lung cancer among non-smokers

  3/3000 = 1 per thousand

RR = 10 / 1 = 10

(lung cancer is 10 times more common among smokers than non smokers)

AR = 10 – 1 / 10 X 100

= 90 %

(90% of the cases of lung cancer among smokers are attributed to their habit of smoking)

Advantages

1) **Incidence** can be directly calculated

2) Direct estimation of the **RR**

3) More than one **outcome** of the RF can be studied

4) **Dose response** relationship can be studied

5) The **temporal association** can be clearly seen in a cohort study

6) **recall bias, interviewer's bias** are not a problem

Disadvantages

1) The major disadvantage is the **time, effort and the money** involved

2) Unsuitable for **rare** diseases

3) **Administrative problems** due to long periods of follow up.

4) **Loss to follow up**- Despite best efforts some loss is inevitable

	Outcome after 10 years	
At the beginning of study	**CHD developed**	**CHD did not develop**
2000 Healthy smokers	65	1,935
4000 Healthy non smokers	20	3,980

Is there any association between smoking and development of Coronary Heart Disease? If yes, justify your result using the table above.

	Disease (Respiratory disease)	Not disease (no respiratory disease)	Total
Exposed to gasoline	60	140	200
Not exposed to gasoline	25	75	100
Total	85	215	300

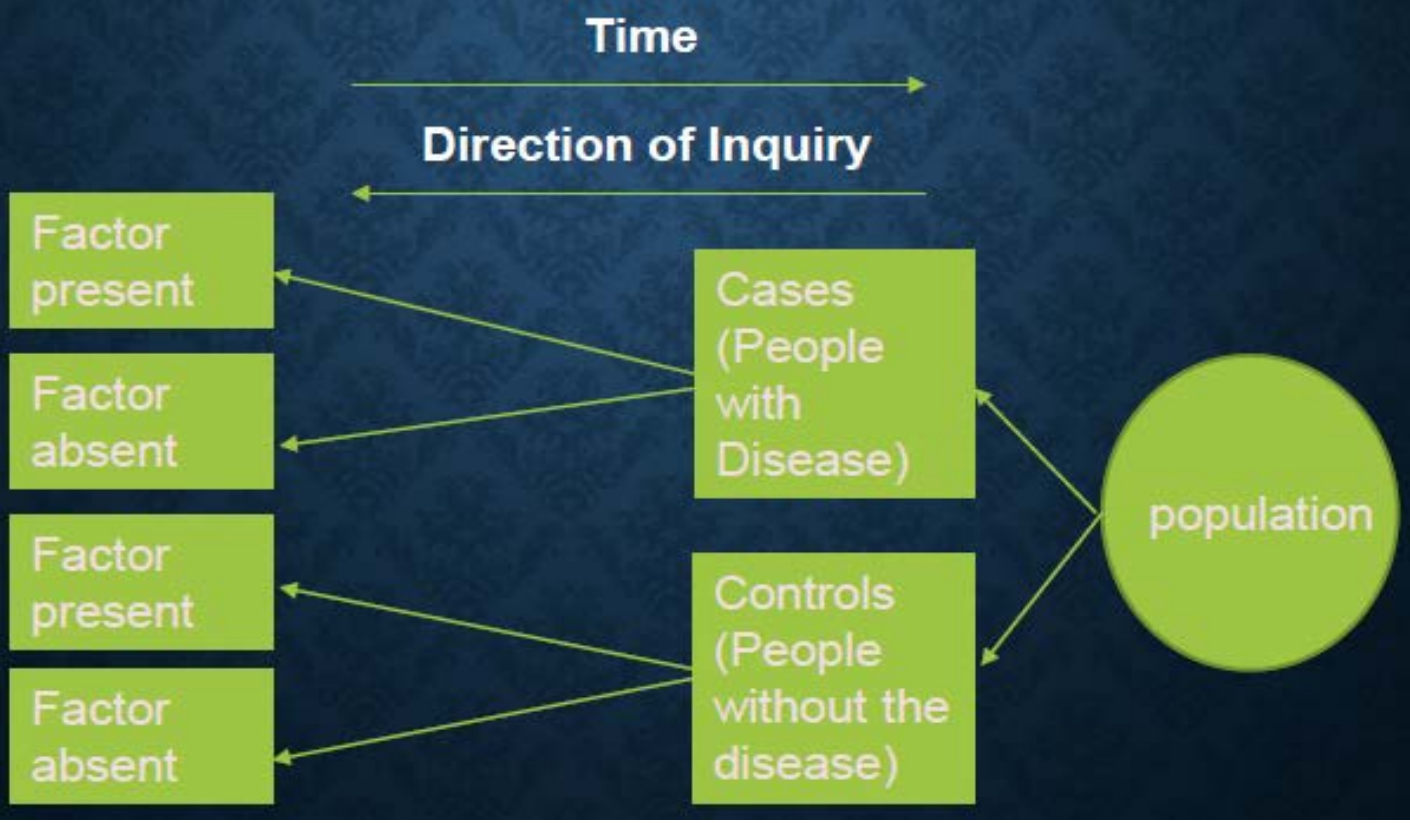
Is there any association between gasoline and respiratory disease..

Exercise

Exposure	Lung Cancer	No lung Cancer	
smokers	70	6930	7000
Non Smokers	3	2997	3000
Total	73	9927	10,000

Case control studies

DESIGN OF CASE CONTROL STUDY



DISTINCT FEATURES

A. both exposure and outcome (disease) have occurred before the start of the study

B. the study proceeds backwards from effect to cause; and

C. it uses a control or comparison group to support or refute an inference

By definition, a case control study involves two populations - cases and controls.

In case control studies, the unit is the INDIVIDUAL rather than the group.

The focus is on a disease or some other health problem that has ALREADY DEVELOPED.

CASE CONTROL STUDY

2 x 2 table which provides a framework

To illustrate, if it is our intention to test the hypothesis that "cigarette smoking causes lung cancer", using the case control method, the investigation begins by assembling a group of lung cancer cases (a + c), and a group of suitably matched controls (b + d).

One then explores the past history of these two groups for the presence or absence of smoking, which is suspected to be related to the occurrence of cancer lung.

If the frequency of smoking, *a/(a* +c) is higher in cases than in controls *b/(b+d),* an association is said to exist between smoking and lung cancer.

Case control studies have their major use in the chronic disease problem when the causal pathway may span many decades.

Basic Design Of A Case Control Study

Suspected or Risk Factor	Cases (Disease present)	Control (Disease absent)
Present	a	B
Absent	c	d
	a+c	b+d

CASE CONTROL STUDY

The steps of control studies are:

1- Selection of cases and identification of source of controls

2- Matching and selection of controls

3- Measurement of exposure in the cases and controls

4- Data analysis

CASE CONTROL STUDY

Step-2

Matching and SELECTION OF CONTROLS

Matching is defined as the process by which we select controls in such a way that they are similar to cases with regard to certain pertinent selected variables (e.g., age) which are known to influence the outcome of disease and which, if not adequately matched for comparability, could distort or confound the results.

'Matching' done to ensure comparability of cases and controls with regard to the 'Confounding factors'.

- **Confounding factor** is one that is **independently associated with the disease** as well as **with the risk factor** and is **distributed unequally** in both the groups.

Advantages of Case Control Studies

1) Easy to conduct
2) Require comparatively few subjects
3) No attrition
4) Gives faster results
5) Inexpensive
6) Suited especially for diseases those are **rare** or **newly identified**.
7) Suitable for diseases which have a long latency period
8) More than one RF's for the disease can be studied simultaneously.
9) Minimal ethical problems

Disadvantages of Case control studies

1) Study depends upon the subject's memory

2) Finding an appropriate control may be difficult

3) We can only establish an association between the disease and the RF. We can't say what came first.

4) Can't calculate the Relative Risk.

5) Case control design may not be suitable for the purpose

Interpretation of Odds Ratio

If OR = 1, the exposure is not related to disease

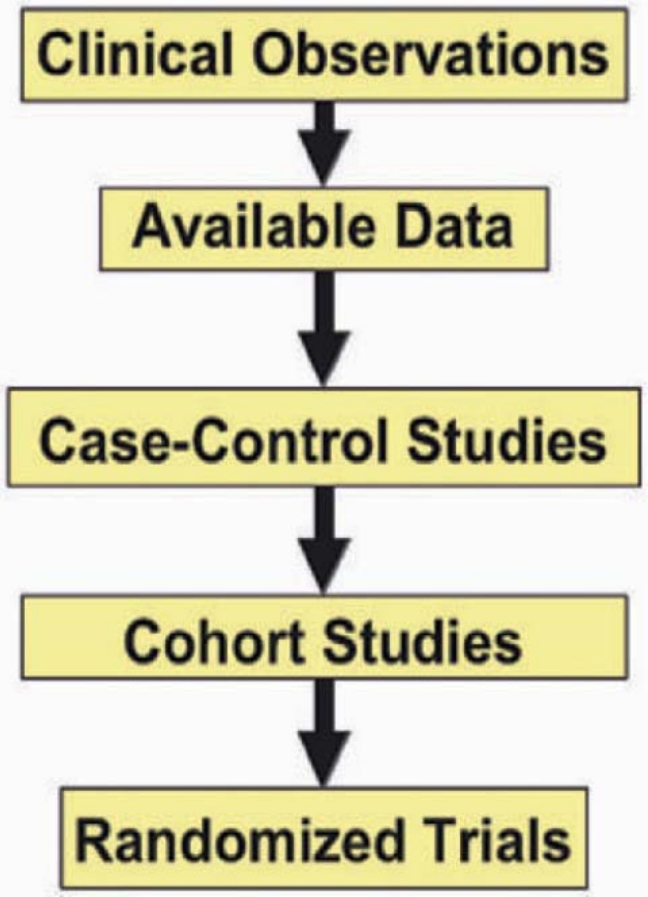
If OR> 1, the exposure is positively related to disease

If OR< 1, the exposure is negatively related to disease

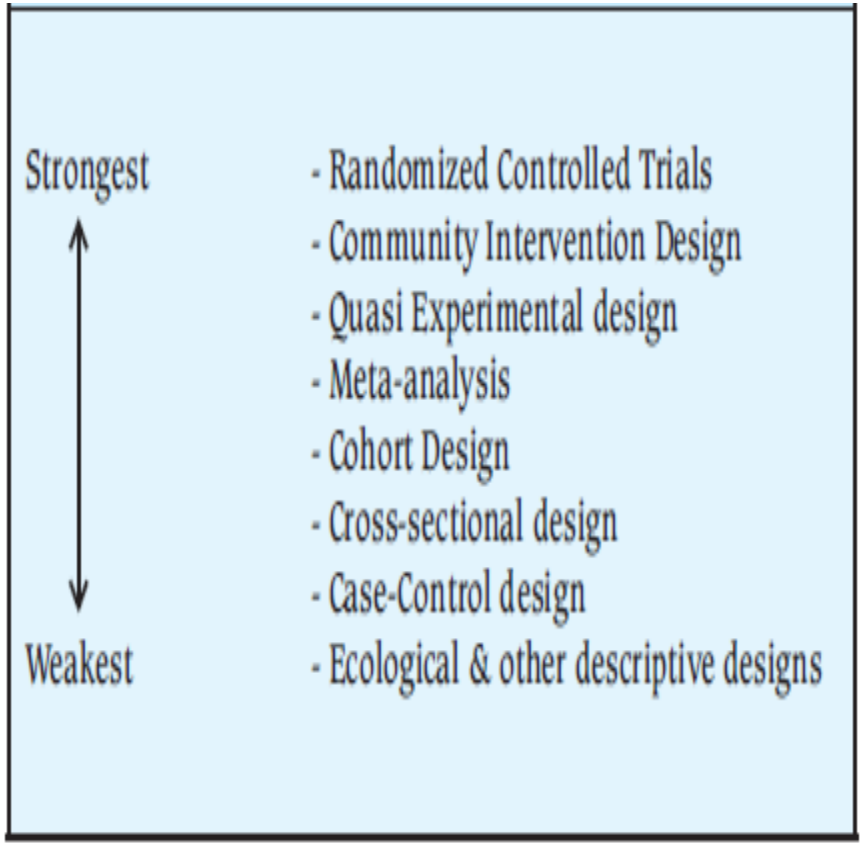
Association and Causation

Guidelines of Causality

APPROACHES TO ETIOLOGY IN HUMAN POPULATIONS



A frequent sequence of studies in human populations



Strength of evidence based on type of epidemiological design

Two step process to carry out studies and evaluate evidence

1. **Determine if an association is present**
   - Ecologic studies: studies of group characteristics
   - Cross-sectional studies: studies at one particular time
   - Case-control or cohort studies: studies of individual characteristics.
2. If an association is demonstrated, **determine whether the observed association is likely to be a causal one** using pre-determined criteria.

Nine guidelines for judging whether an association is causal

- Temporal relationship
- Strength of association
- Dose response relationship
- Replication of the findings
- Biologic plausibility
- Consideration of alternate explanations
- Cessation of exposure
- Specificity of the association
- Consistency with other knowledge

Temporal Relationship

- Exposure to the factor must have occurred before the disease developed.
- Easiest to establish in a cohort study
- Length of interval between exposure and disease very important
  - If the disease develops in a period of time too soon after exposure, the causal relationship is called into question

Strength of Association

- The larger the **RELATIVE RISK OR ODDS RATIO**, the higher the likelihood that the relationship is causal

- However, care must be taken to examine confidence intervals and sample size
  - For example, if the confidence interval is wide (e.g., 1.8 - 22.6), an OR of 12.0 is less strong because we are less confident of the strength of the odds ratio

Strength of association

Which odds ratio would you be more likely to infer causation from?

OR#1:	OR = 1.4	95% CI = (1.2 - 1.7)
OR#2:	OR = 9.8	95% CI = (1.8 - 12.3)
OR#3:	OR = 6.6	95% CI = (5.9 - 8.1)

Dose-Response Relationship

- With increasing dose, there is increasing risk of disease

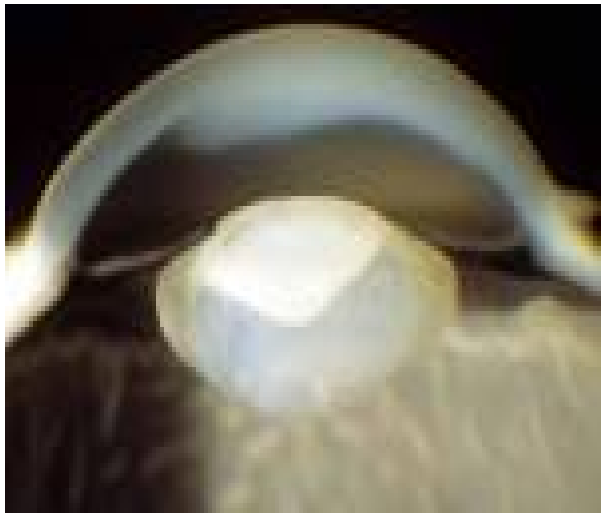
- This is not considered necessary for a causal relationship, but does provide additional evidence that a causal relationship exists

Replication of the Findings

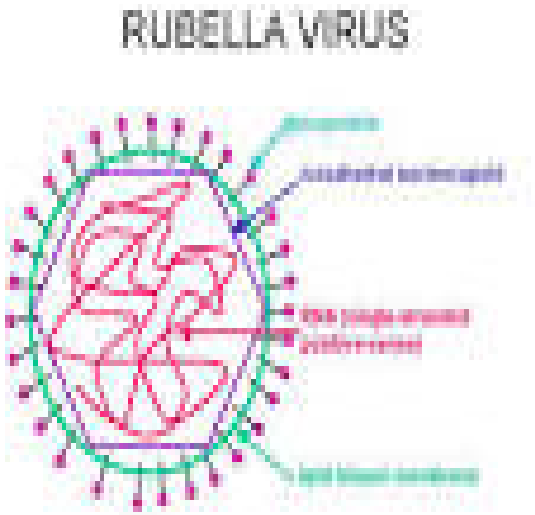
- If there is a true causal relationship between exposure and disease, the expectation is that we would see the association consistently in other (NOT necessarily all) subgroups of the population

Biologic plausibility

- Consistency of epidemiologic plausibility with existing biologic knowledge
- Requires knowledge of the biologic etiology of disease
- Gregg's observations on rubella and congenital cataracts preceded any knowledge of teratogenic viruses



Congenital Cataract



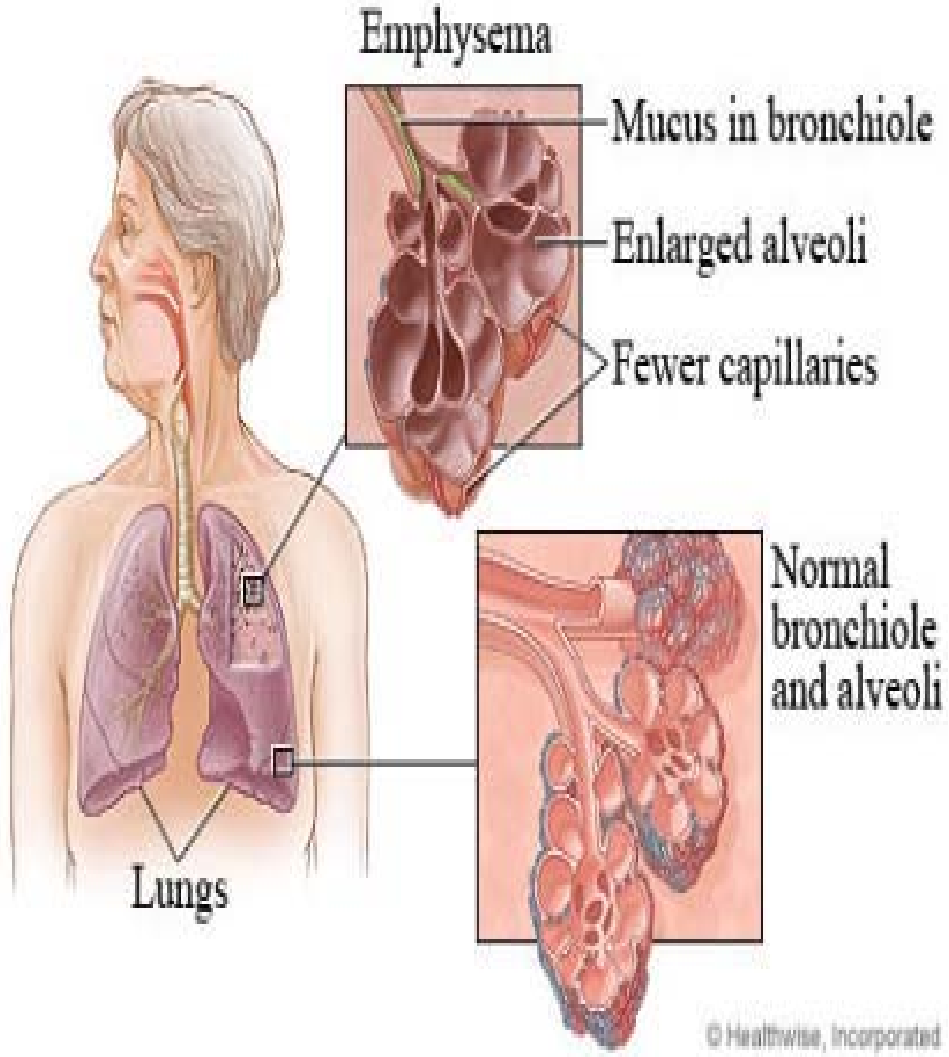
Rubella Virus

Consideration of alternate explanations

- Requires a knowledge of the literature and known risk factors for the disease

Cessation of exposure

- Upon elimination or reduction of exposure to the factor, the risk of disease declines.
- HOWEVER, in certain cases, the damage may be irreversible.
- Example: Emphysema is not reversed with the cessation of smoking, but its progression is reduced.



Specificity of the Association

- The weakest of the criteria (should probably be eliminated)
- Specific exposure is associated with only one disease
- This is used by tobacco companies to argue that smoking is not causal in lung cancer
  - Smoking is associated with many diseases
- If anything, may provide support for a causal relationship, but does not negate if not present

Diseases screening

Criteria for Screening

The Disease	Serious, high prevalence of preclinical stage, natural history understood, long period b/w 1st signs & overt signs
The Population	Identification of risk groups, attitudes to screening
The Test	Sensitive & specific, reliable, simple & affordable, safe & acceptable
The Treatment	Effective, acceptable, and safe treatment available
The Evaluation	The cost of the programme, screening participation rate

Screening Test vs. Disease State: Four Possible Scenarios

Test	**Disease state**		TOTAL
	Diseased	Disease-free	
Positive	D^+T^+	D^-T^+	Tested positive
Negative	D^+T^-	D^-T^-	Tested negative
TOTAL	Total diseased	Total disease-free	All subjects

Test	Disease State		TOTAL
	Diseased	Disease-free	
+ve	a (True Positive [TP])	b (False Positive [FP])	a+b
-ve	c (False Negative [FN])	d (True Negative [TN])	c+d
TOTAL	a+c	b+d	a+b+c+d

Sensitivity: proportion of individuals with the disease who are correctly identified by the test = TP/(TP+FN) = a/(a+c)

Specificity: proportion of individuals without the disease who are correctly identified by the test = TN/(TN+ FP) d/(b+d)

Test	Disease State		TOTAL
	Diseased	Disease-free	
+ve	a (True Positive [TP])	b (False Positive [FP])	a+b
-ve	c (False Negative [FN])	d (True Negative [TN])	c+d
TOTAL	a+c	b+d	a+b+c+d

Positive Predictive Value: Proportion of individuals with a positive test result who have the disease = TP/(TP+FP) = a/(a+b)

Negative Predictive Value : Proportion of individuals with a negative test result who do not have the disease = TN/(TN+FN) = d/(c+d)

Positive & Negative Predictive Value

Test Result	Disease		TOTAL
	Present	Absent	
+ve	350	1900	2250
-ve	150	7600	7750
TOTAL	500	9500	10,000

Positive Predictive Value: Proportion of individuals with a positive test result who have the disease = TP/(TP+FP)= a/(a+b) 350/2250=15.6%

Negative Predictive Value : Proportion of individuals with a negative test result who do not have the disease = TN/(TN+FN) = d/(c+d) 150/7750=98.1%

Exercise

Screening test	Disease state		Total
	Diseased	Not diseased	
Positive	20	60	80
Negative	80	40	120
	100	100	200