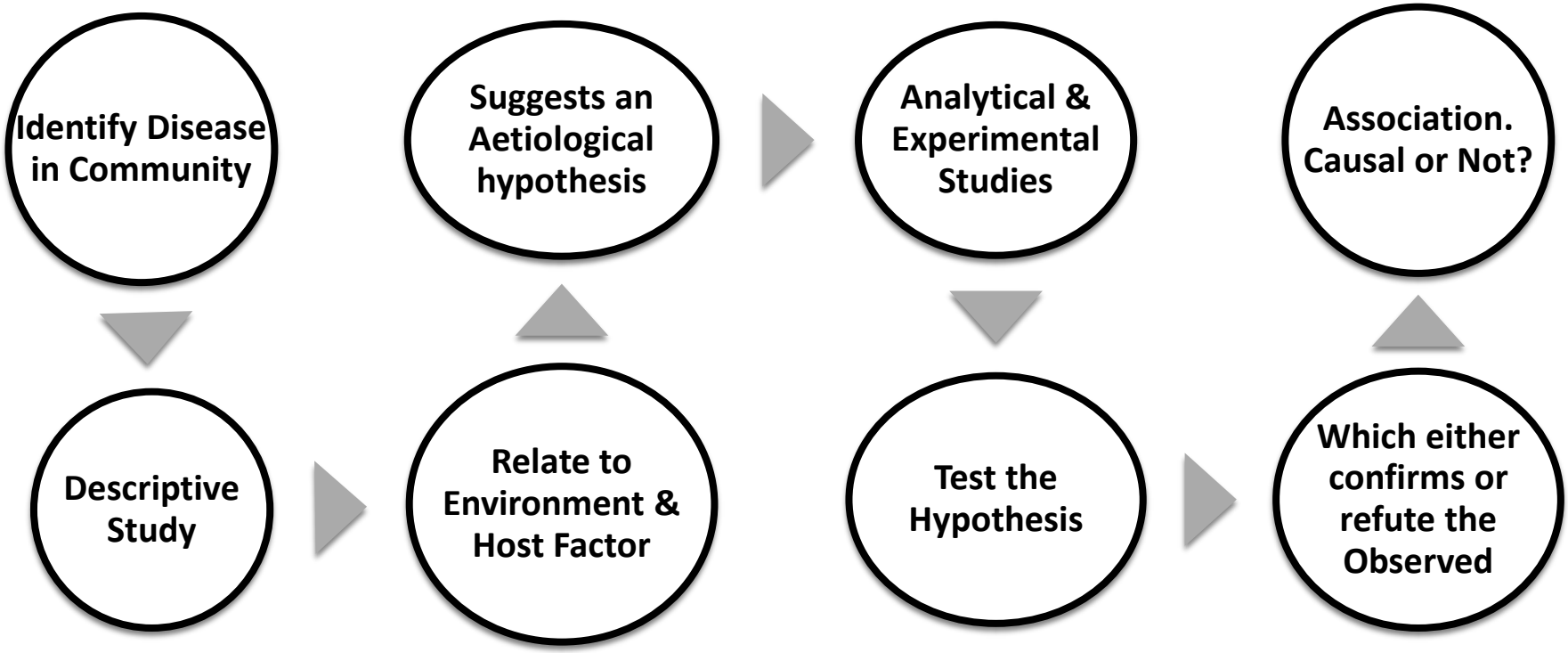
Association & Causation

Dr Farzana Islam

Introduction:

- Epidemiological principles stand on two basic assumptions
  - Human disease **does not occur at random**
  - The disease and its **causal as well as preventive factors** can be identified by a thorough investigation of population
- Hence, identification of causal relationship between a disease and suspected risk factors forms part of epidemiological research.

Introduction:…….contd



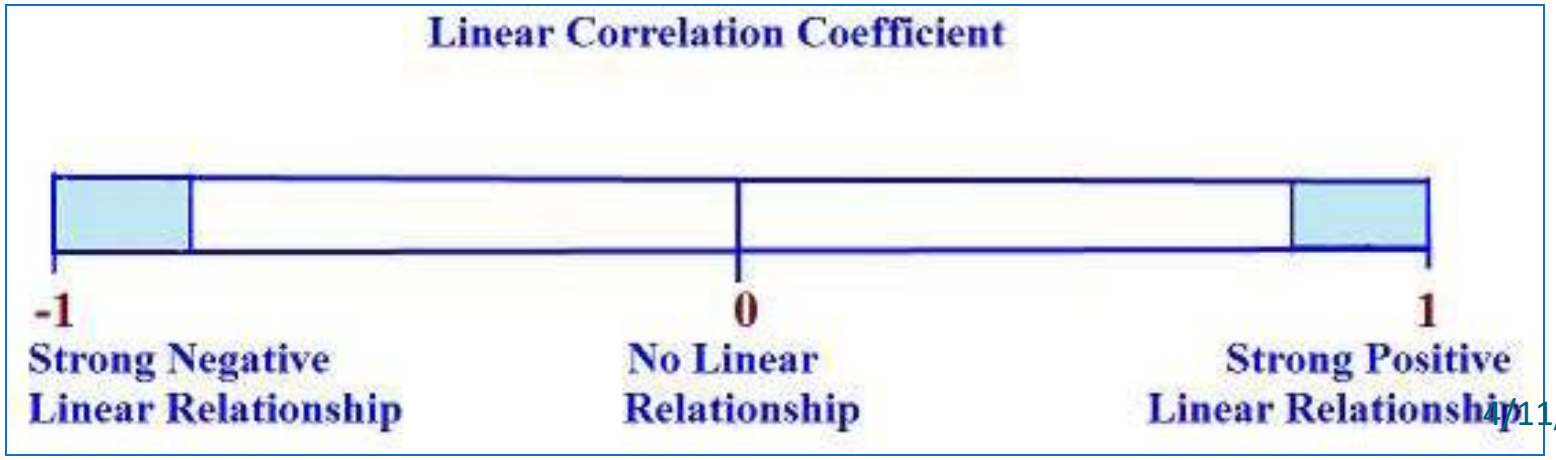
Association:

- Defined as the occurrence of two variables more often than would be expected by chance.

  - **Association does not necessarily mean causation relationship**

Correlation:

- Defined as the ***degree of association*** between the two characteristics.
- Correlation co-efficent ranges from -1.0 to + 1.0.
  - Value of 1.0 means that the two variables exhibit a perfect linear relationship.
- Causation implies correlation, but correlation does not imply causation



Correlation:

- Defined as the ***degree of association*** between the two characteristics.
- Correlation co-efficent ranges from -1.0 to + 1.0.
  - Value of 1.0 means that the two variables exhibit a perfect linear relationship.
- **Causation implies correlation, but correlation does not imply causation**

Causation

- **CAUSE - an event, condition, characteristic (or a combination)** which plays an important role in the occurrence of the outcome (e.g. smoking and lung cancer)
- **Factors involved in Causation**
  - **Precipitating Factors, e.g.** Exposure to agent
  - **Predisposing Factors, e.g.** Age, Sex, Previous Illness
  - **Enabling Factors e.g.** Low SEC, malnutrition
  - **Reinforcing Factors e.g.** Repeated Exposure

Types Of Association

- Spurious association
- Indirect association
- Direct association
  - One-to-one causal association
  - Multifactorial causation

Spurious Association

- This is an association which appears due to improper comparison.
- Observed association between a disease and suspected factor may not be real.
  - E.g.; Neonatal mortality was observed to be more in the newborns born in a hospital than those born at home. This is likely to lead to a conclusion that home delivery is better for the health of newborn.
  - However, this conclusion was not drawn in the study because the proportion of “high risk” deliveries was found to be higher in the hospital than in home.

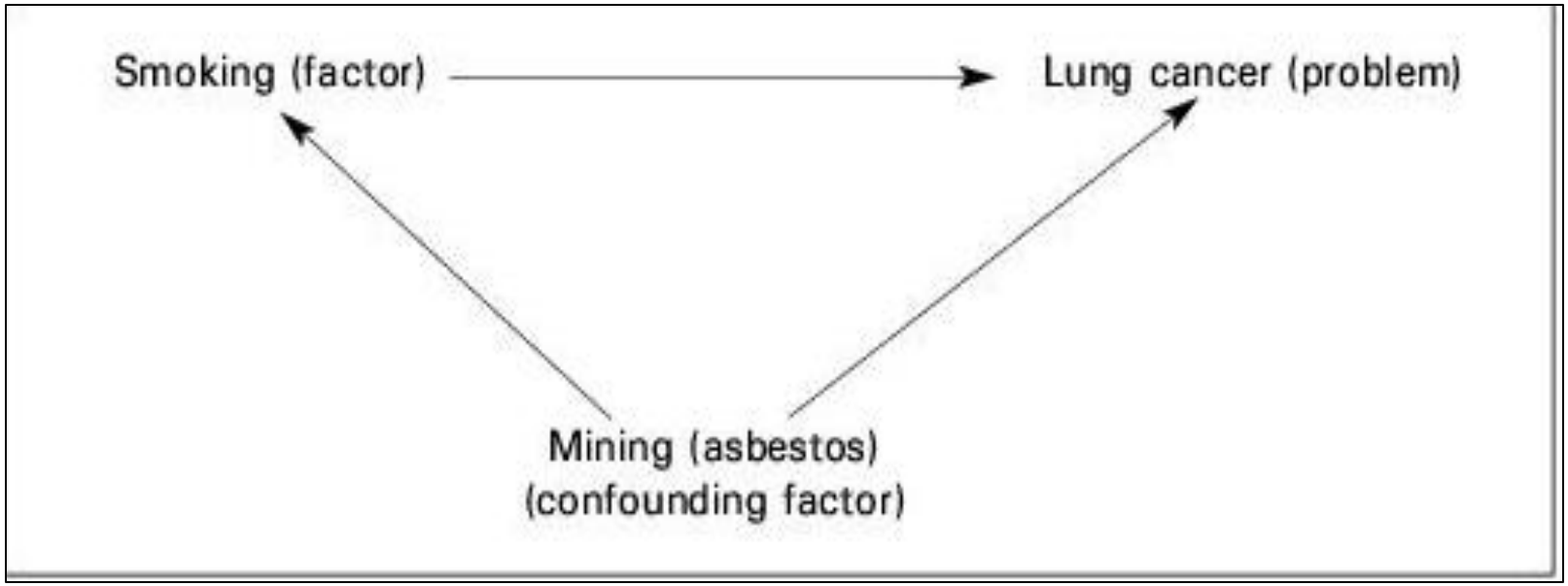
Indirect Association

- It is a statistical association between a characteristic of interest and a disease due to the presence of another factor i.e. common factor (confounding variable).
  - E.g.: **Neonatal mortality (A)** was found to be associated with **maternal age** above 30 years **(B)** and with **birth order 4** and above **(C).**
  - It was also shown that the attribute B and C were associated with each other.

What is a Confounding Variable?

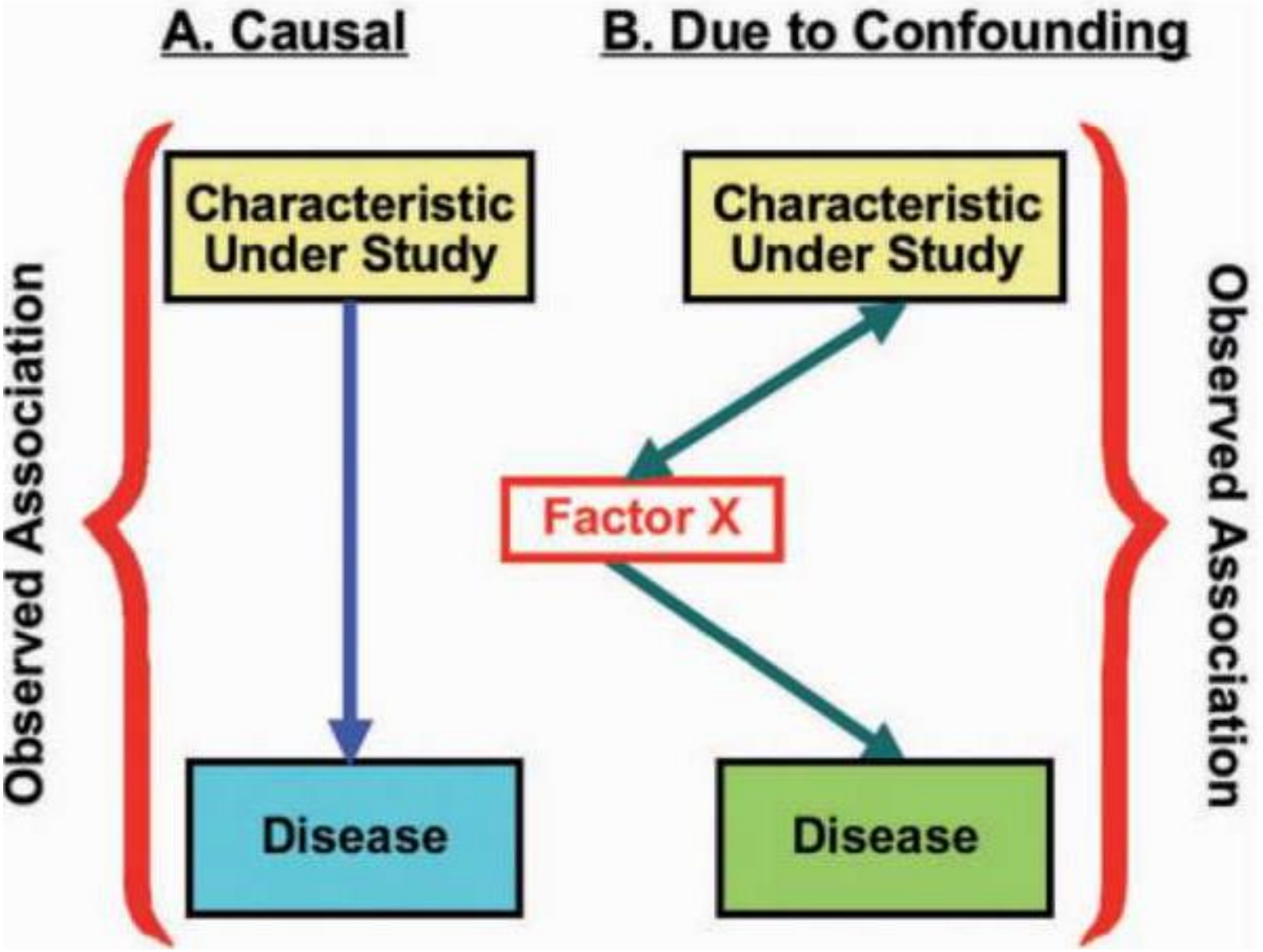
- A confounding variable is an “extra” variable that was not accounted for.
- They can ruin an experiment and give useless results.
- They can suggest there is correlation when in fact there isn’t.
- They can even introduce bias.
- That’s why it’s important to know what one is, and how to avoid getting them into your experiment in the first place.
- A confounding variable can have a hidden effect on experiment’s outcome.

Examples of Confounding Variables:



Direct Association

- The association between the two attributes is not through the third attributes.
- When the study reveals it is not a spurious association.
- When the disease is present, the factor must also be present.


A. Causal
B. Due to Confounding
Characteristic Under Study
Characteristic Under Study
Factor X
Disease
Disease
Observed Association
Observed Association

Direct Association Is Classified Into Two Types

- **One-to-one Casual Relationship**
- **Multifactorial causation:**

One-to-one Casual Relationship

- The variables are stated to be casual related (AB) if a change in A is followed by a change in B.
- When the disease is present, the factor must also be present.
- A single factor or cause may lead to more than one outcome.

Multifactorial causation:

- Alternative causal factors each acting independently.
- E.g. In lung cancer more than one factor (e.g. air pollution, smoking, heredity) can produce the disease independently.

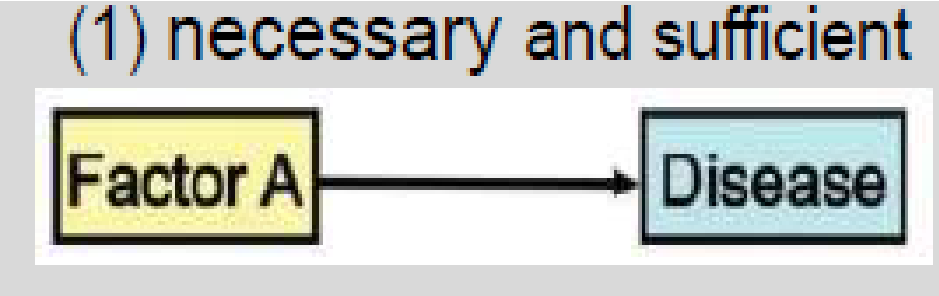
TYPES OF CAUSAL RELATIONSHIPS

If a relationship is causal, four types of causal relationships are possible:

1. necessary and sufficient
2. necessary, but not sufficient
3. sufficient, but not necessary
4. neither sufficient nor necessary

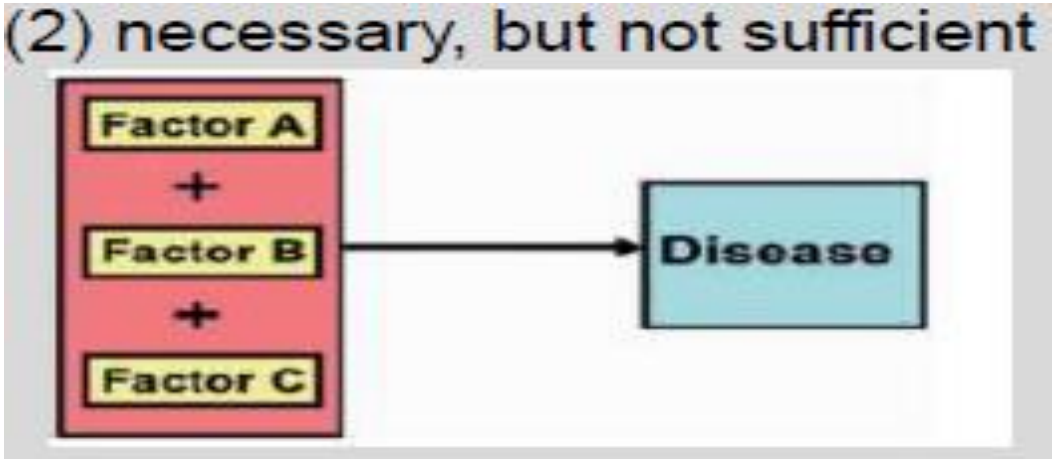
1. Necessary and Sufficient

- A Factor is both necessary and sufficient for producing the disease.
- Without that factor, the disease never develops (the factor is necessary), and in the presence of that factor, the disease always develops (the factor is sufficient).



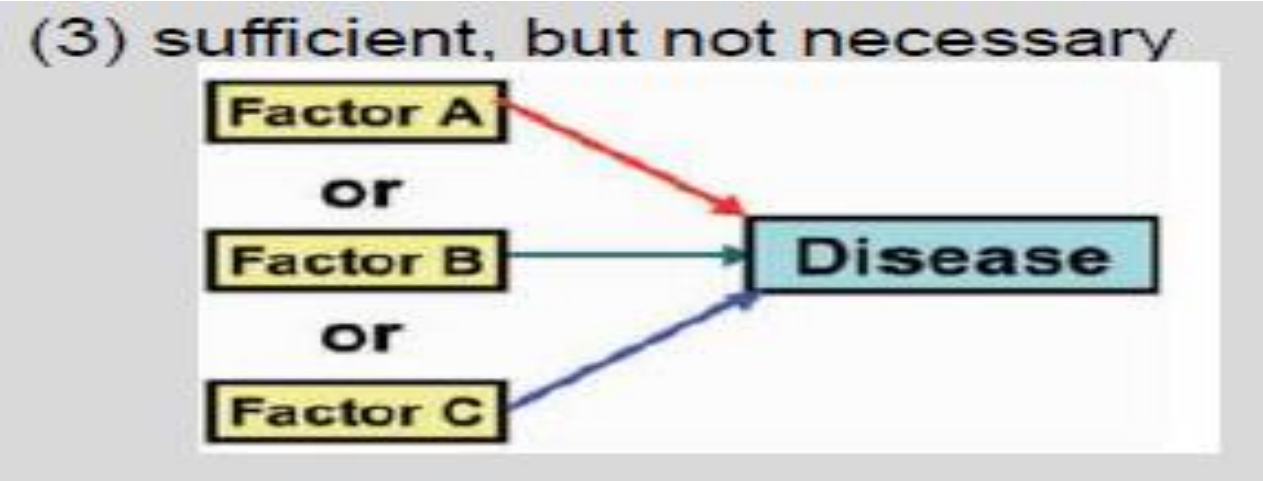
2. Necessary But not Sufficient

- Each factor is necessary, but not, in itself, sufficient to cause the disease
- For example, carcinogenesis is considered to be a multistage process involving both initiation and promotion.
- For cancer to result, a promoter must act after an initiator has acted.
- Action of an initiator or a promoter alone will not produce a cancer.



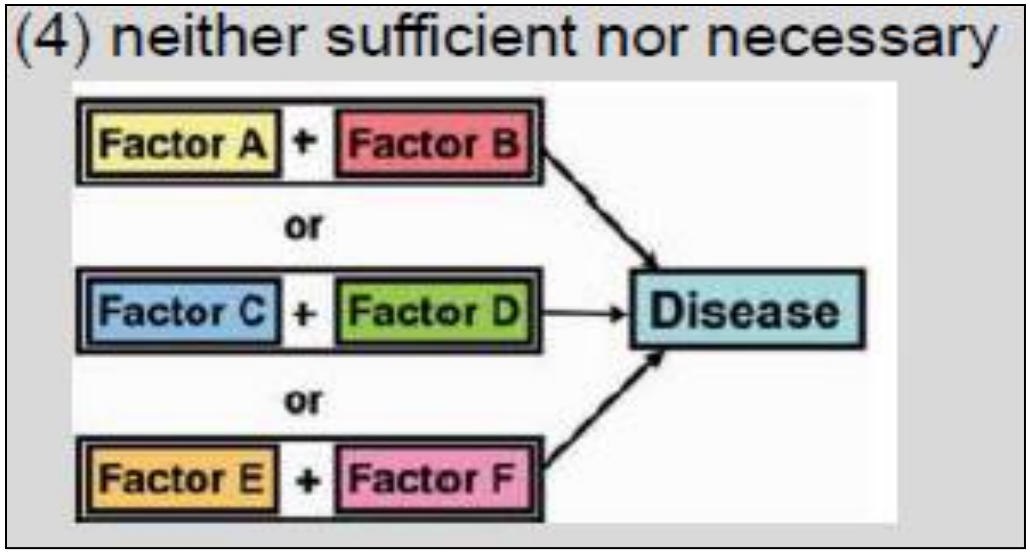
3. Sufficient But Not Necessary

- The factor alone can produce the disease, but so can other factors that are acting alone
- For example, either radiation exposure or benzene exposure can each produce leukemia without the presence of the other.



4. Neither Sufficient Nor Necessary

- Here the factor, by itself, is neither sufficient nor necessary to produce disease
- This is a more complex model, which probably most accurately represents the causal relationships that operate in most chronic diseases



Criteria for Causal Association

Hill's Criteria (1965)	Surgeon General's Report (1964)
• 1. Strength • 2. Consistency • 3. Specificity • 4. Temporality • 5. Biological gradient • 6. Plausibility • 7. Coherence • 8. Experiment • 9. Analogy	• 1. Consistency • 2. Strength —Dose-response • 3. Specificity • 4. Temporality • 5. Coherence

Criteria for Causal Association….contd

Criteria for Causal Association (Gordis2014)

- 1.Temporal relationship
- 2.Strength of the association
- 3. Dose-Response Relationship
- 4. Replication of the Findings
- 5. Biologic Plausibility
- 6. Consideration of Alternate Explanations
- 7. Cessation of Exposure
- 8. Consistency with Other Knowledge
- 9. Specificity of the Association.

1.Temporal relationship

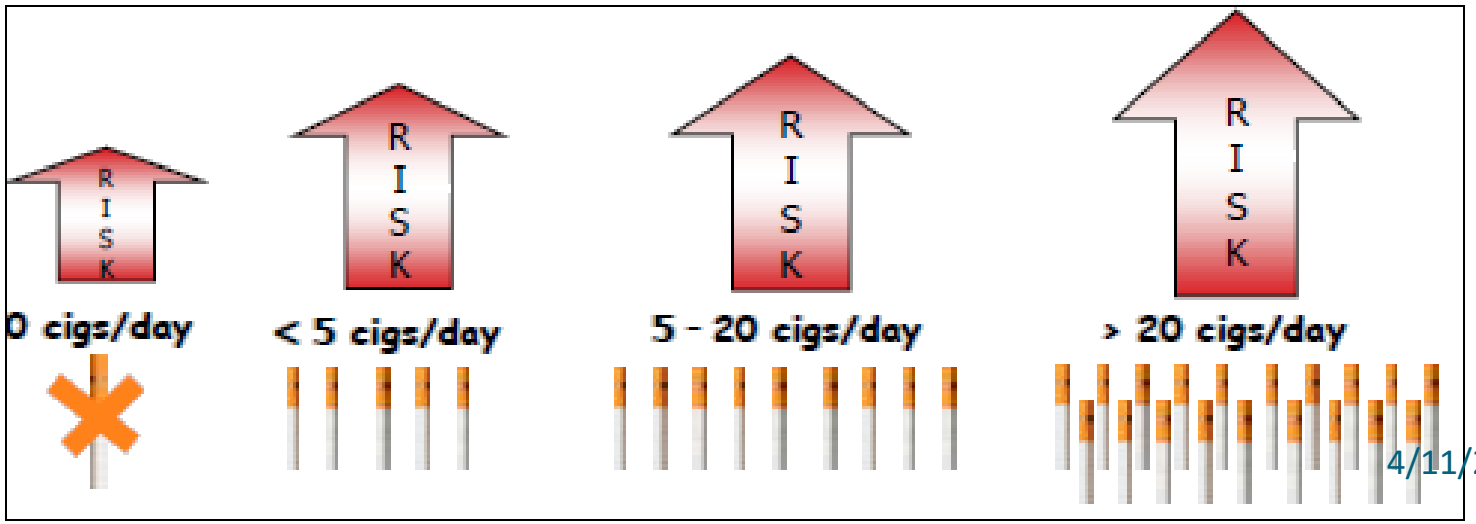
- The causal attribute must precede the disease or unfavorable outcome.
- Exposure to the factor must have occurred before the disease developed.
- Length of interval between exposure and disease very important
- If the disease develops in a period of time too soon after exposure, the causal relationship is called into question

2.Strength of the association

- Relationship between cause and outcome could be strong or weak.
- With increasing level of exposure to the risk factor an increase in incidence of the disease is found.
- There are statistical methods to quantify the strength of association viz; calculation of relative risk, attributable risk etc.

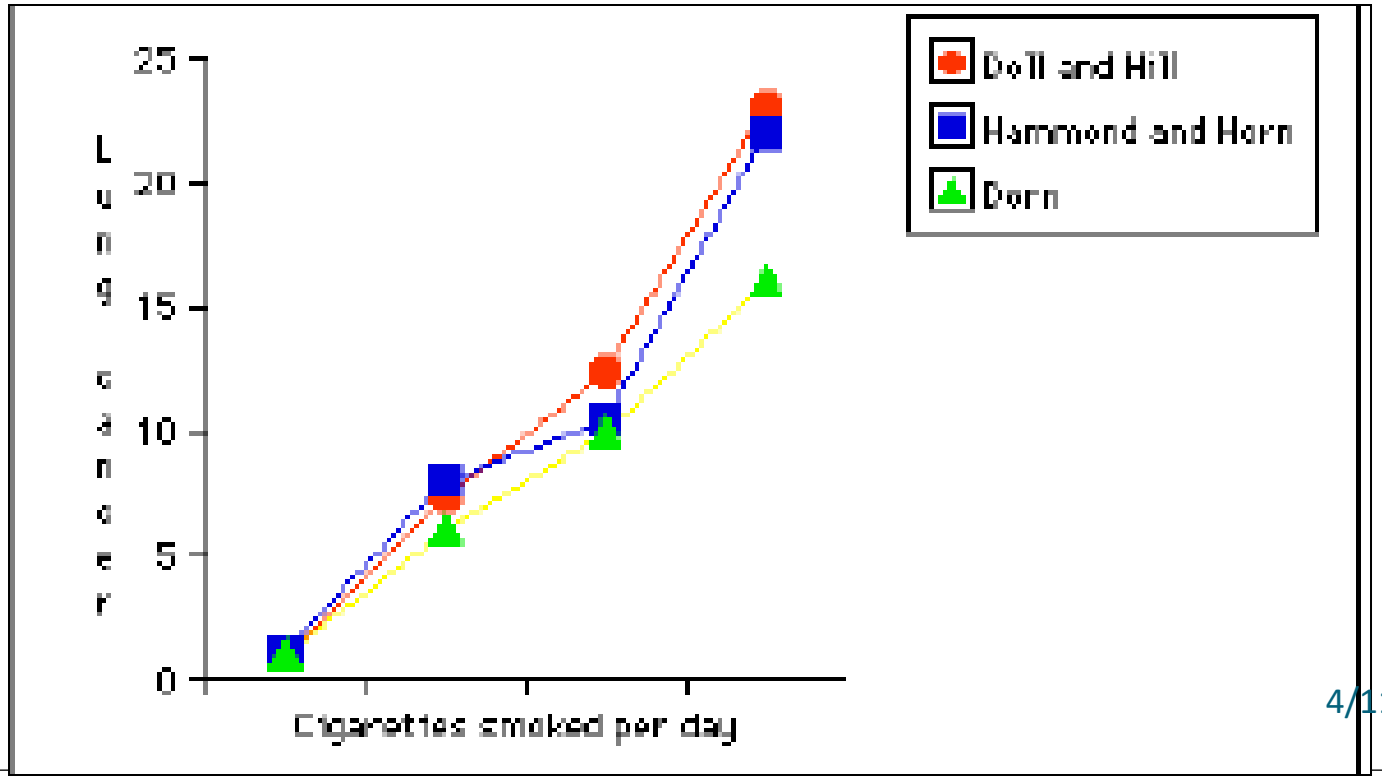
3.Dose-Response Relationship

- As the dose of exposure increases, the risk of disease also increases
- If a dose-response relationship is present, it is strong evidence for a causal relationship.
- E.g. death rate from lung cancer was increased by number of cigarettes smoking (British doctors 1951-1981)



4.Replication of the Findings

- If the relationship is causal, we would expect to find it consistently in different studies and in different populations
- Replication of findings is particularly important in epidemiology.
- If an association is observed, we would also expect it to be seen consistently within subgroups of the population and in different populations.

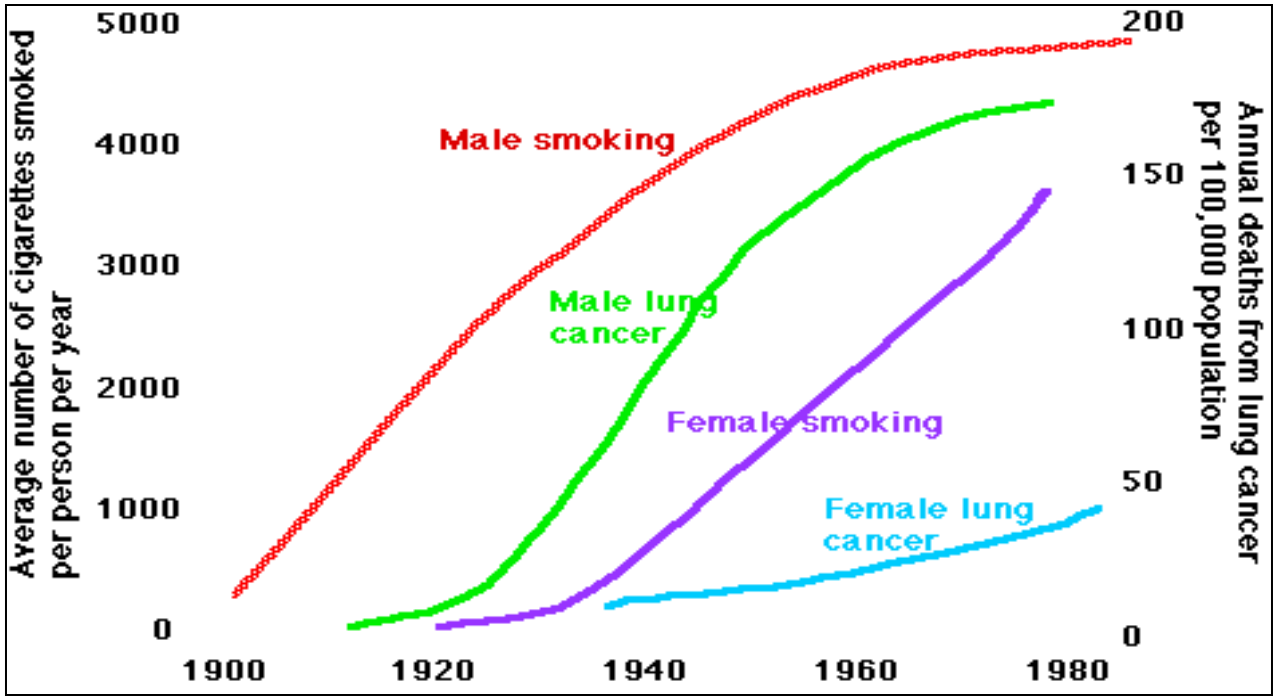


5.Biologic Plausibility

- The association must be consistent with the current knowledge of disease. (viz mechanism of action, evidence experiments etc).
- Sometimes the lack of plausibility may simply be due to the lack of sufficient knowledge about the pathogenesis of a disease

6. Coherence of the Association

- The association must be coherent with the known facts of relevant origins.
- Male and Female differences in trends of lung cancer deaths is coherent with recent adoption of Cigarette smoking by women

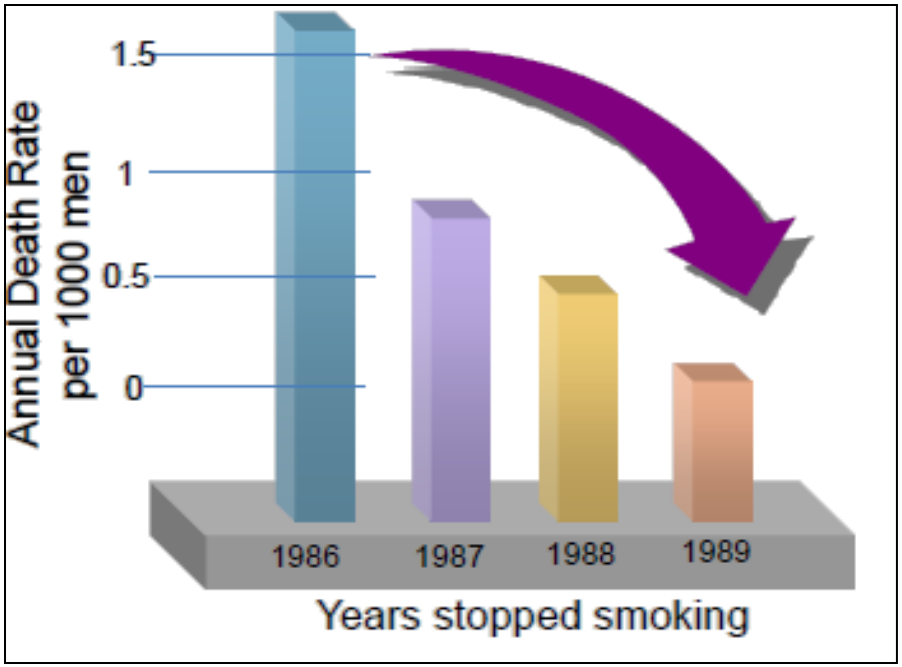


7. Consideration of Alternate Explanations

- In judging whether a reported association is causal, the extent to which the investigators have taken other possible explanations into account and the extent to which they have ruled out such explanations are important considerations.
- Deriving Causal inferences by eliminating- Bias, Confounding and Chance etc

8.*Cessation of Exposure*

- If a factor is a cause of a disease, we would expect the risk of the disease to decline when exposure to the factor is reduced or eliminated



9. *Specificity of the Association*

- Specificity implies a one to one relationship between the cause and effect (Weakest Criteria).
  - Not everyone who smokes develop Lung Cancer,
  - Not everyone who develops cancer has smoked.
- Lack of specificity does not negate causation.

Measures of association

- Ratio measures:
  - measures of association in which *relative differences* between groups being compared
- Difference measures:
  - Difference measures are measures of association in which *absolute differences between groups being compared .*

Absolute differences

- Main goal is often an absolute reduction in the risk of an undesirable outcome.
- When outcome of interest is continuous, the assessment of mean absolute differences between exposed and unexposed individuals may be an appropriate method for the determination of association

Relative differences

- Can be assessed for discrete outcomes.
- To assess causal associations

Type	Example	Usual application
Absolute difference	AR (Attributable Risk)	Primary prevention impact: search for causes.
	PAR(Population Attributable risk)	Primary prevention impact
	Efficacy	Impact of intervention on recurrences, case fatality etc.
	Mean differences (continuous outcome)	Search for determinants
Relative difference	Relative risk/rate	Search for causes
	Relative odds (ODDS ratio)	Search for causes

$$\text{Relative risk} = \frac{\text{Incidence among exposed}}{\text{Incidence among unexposed}}$$

It is direct measure of the strength of association.

RR = 1	No association
RR > 1	Positive association (possibly causal)
RR < 1	Negative association (possibly protective)

Odd's ratio(Relative odds)

	Cases(with disease)	Controls (without disease)
H/O of expos	a	b
No H/O expos	c	d

product of the two cells that support the hypothesis of an association : product of the two cells that negate the hypothesis of an association

ad : bc

ATTRIBUTABLE RISK (AR)

- How much of the disease that occurs can be attributed to a certain exposure?
- It is often used to imply a cause-effect relationship and should be interpreted as a true etiologic fraction only when there is a reasonable certainty of a causal connection between exposure and outcome
- When causality has not been firmly established then the AR is termed as excess fraction

POPULATION ATTRIBUTABLE RISK

- What proportion of the disease incidence in a total population can be attributed to a specific exposure?
- To know the PAR , we need to know
  - incidence in total population =a
  - incidence in unexposed group(background risk)=b
  - PAR= a-b ÷ a