

# ASSOCIATION & CAUSATION IN EPIDEMIOLOGICAL STUDIES

**Dr. Sireen Alkhalidi**

**Community Medicine 2018/ 2019**

**The University of Jordan**

# Which of these foods will stop cancer? (Not so fast)

- Cancer patients always ask what to eat to reduce their chances of dying from the disease.
- Diet messages are everywhere:
  - NCI: Eat 5 to 9 fruits and vegetables a Day for Better Health
  - Prostate Cancer Foundation has anticancer diet
- Will dietary changes make a difference?
- It is more difficult than expected to discover if diet affects cancer risk.
- Hypotheses are abundant, but convincing evidence remains elusive (hard to prove).

# What is the question?



- Does the exposure lead to an increase (or decreased) risk of disease?
- Is the exposure causal (or protective)?
  - ▣ We observe **associations**
  - ▣ We infer (guess, speculate, reach to a conclusion) about **causes**.

# ASSOCIATION

- **Definition:** the concurrence of two variables more often than would be expected by chance.
- **Types of Associations:**
  1. **Spurious Association:** (Amount of ice cream sold and deaths by drownings) (Shoe size and reading performance for elementary school children)
  2. **Indirect Association**
  3. **Direct (causal) Association**
    1. **One to one causal association**
    2. **Multi-factorial causation.**

# Association and Causation



# Association or not?

---

- A researcher in his observational study found that the average serum homocysteine among patients of IHD was 15 mcg/dl (Normal=10-12 mcg/dl)!

# Implication

- Can we say that
  - ▣ Hyperhomocystenemia causes IHD?
- Hypothesize that
  - ▣ Hyperhomocystenemia **may** have a role in etiology of IHD.
- For final proof there has to be a 'comparison'.
- Comparison would generate another summary measure which shows the extent of 'Association' or 'Effect' or 'risk' (RR, OR, P-value, AR)

# Cause

- Cause defined as “anything producing an effect or a result”. [Webster]
- Cause in medical textbooks discussed under headings like- “**etiology**”, “**Pathogenesis**”, “**Mechanisms**”, “**Risk factors**”.
- Important to physician because it guides their approach to three clinical tasks- **Prevention, Diagnosis & Treatment.**

# Causal Relationships

A causal pathway may be direct or indirect

- In **direct causation**, A causes B without intermediate effects (very rare)
- In **indirect causation**, A causes B, but with intermediate effects

In human biology, intermediate steps are virtually always present in any causal process

# Theories of Disease Causation

- ❑ **Supernatural Theories**
- ❑ **Hippocratic Theory**
- ❑ **Miasma**
- ❑ **Theory of Contagion**
- ❑ **Germ Theory (cause shown via Henle-Koch postulates)**
- ❑ **Classic Epidemiologic Theory**
- ❑ **Multicausality and Webs of Causation (cause shown via Hill's postulates)**

# Hippocratic Theory

Hippocrates promoted the concept that disease was the result of an imbalance among four vital "humors" within us:

- Yellow Bile, Black Bile, Phlegm, Blood

Hippocrates believed that if one of the humors became excessive or deficient, health would deteriorate and symptoms would develop.

Hippocrates was a keen **observer** and tried to relate an individual's **exposures** (e.g., diet, exercise, occupation, and other behaviors) to subsequent **health outcomes**.

# Henle-Koch Postulates (Germ Theory)

Even though there was a "germ" of truth in miasmatic theory, in that it focused attention on environmental causes of disease and partly explained social disparities in health (poor people being more likely to live near foul odors), the theory began to fall into disfavor as the germ theory gained acceptance.

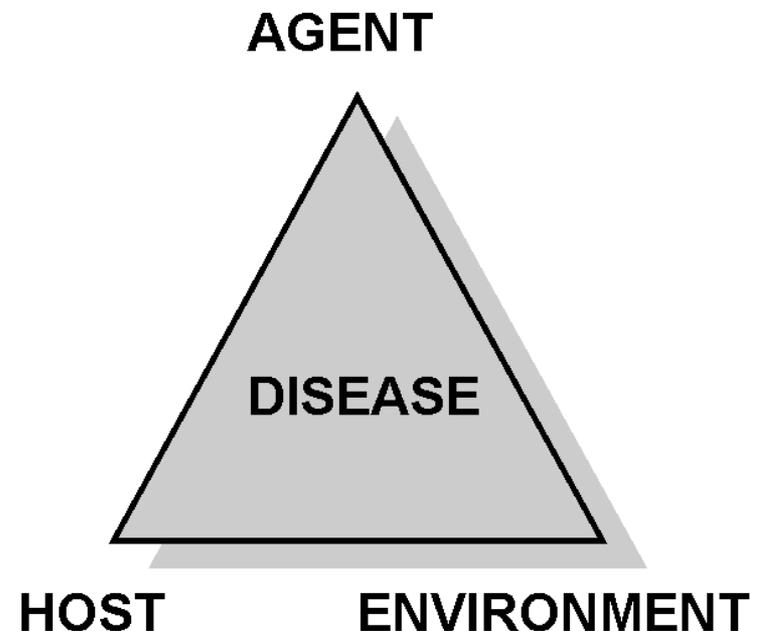
Louis Pasteur and others introduced the germ theory in 1878:

1. The agent is present in every case of the disease
2. It does not occur in any other disease (one agent one disease)
3. It can be isolated and if exposed to healthy subjects will cause the related disease

# Classic Epidemiologic Theory: Epidemiologic Triad

**Disease is the result of forces within a dynamic system consisting of:**

- 1. Agent**
- 2. Host**
- 3. Environment**



# Classic Epidemiologic Theory

- **Agents**

  - Living organisms**

  - Exogenous chemicals**

  - Genetic traits**

  - Psychological factors and stress**

  - Nutritive elements**

  - Endogenous chemicals**

  - Physical forces**

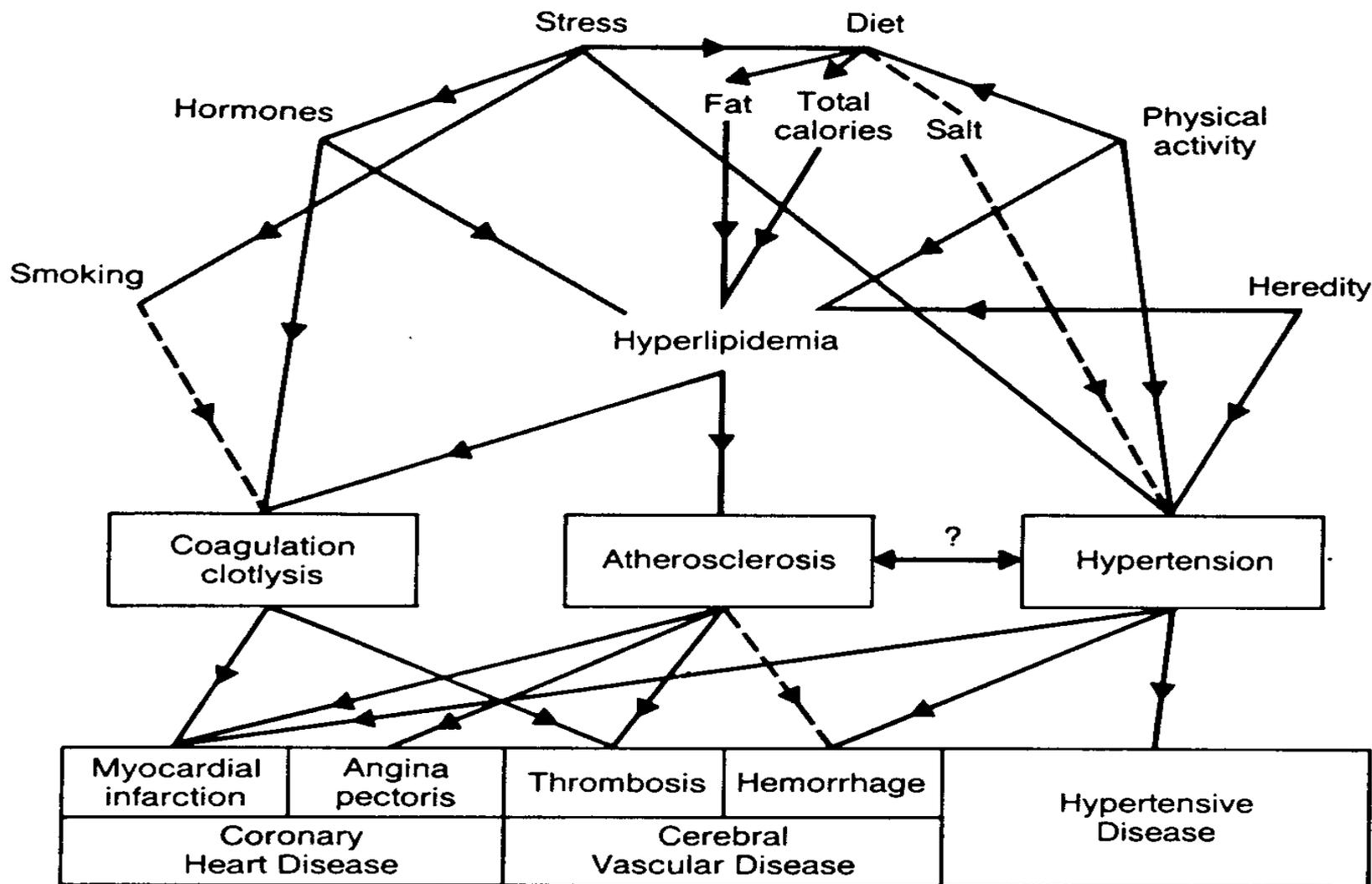
- **Agent factors: characteristics such as infectivity, pathogenicity and virulence (ability to cause serious disease)**

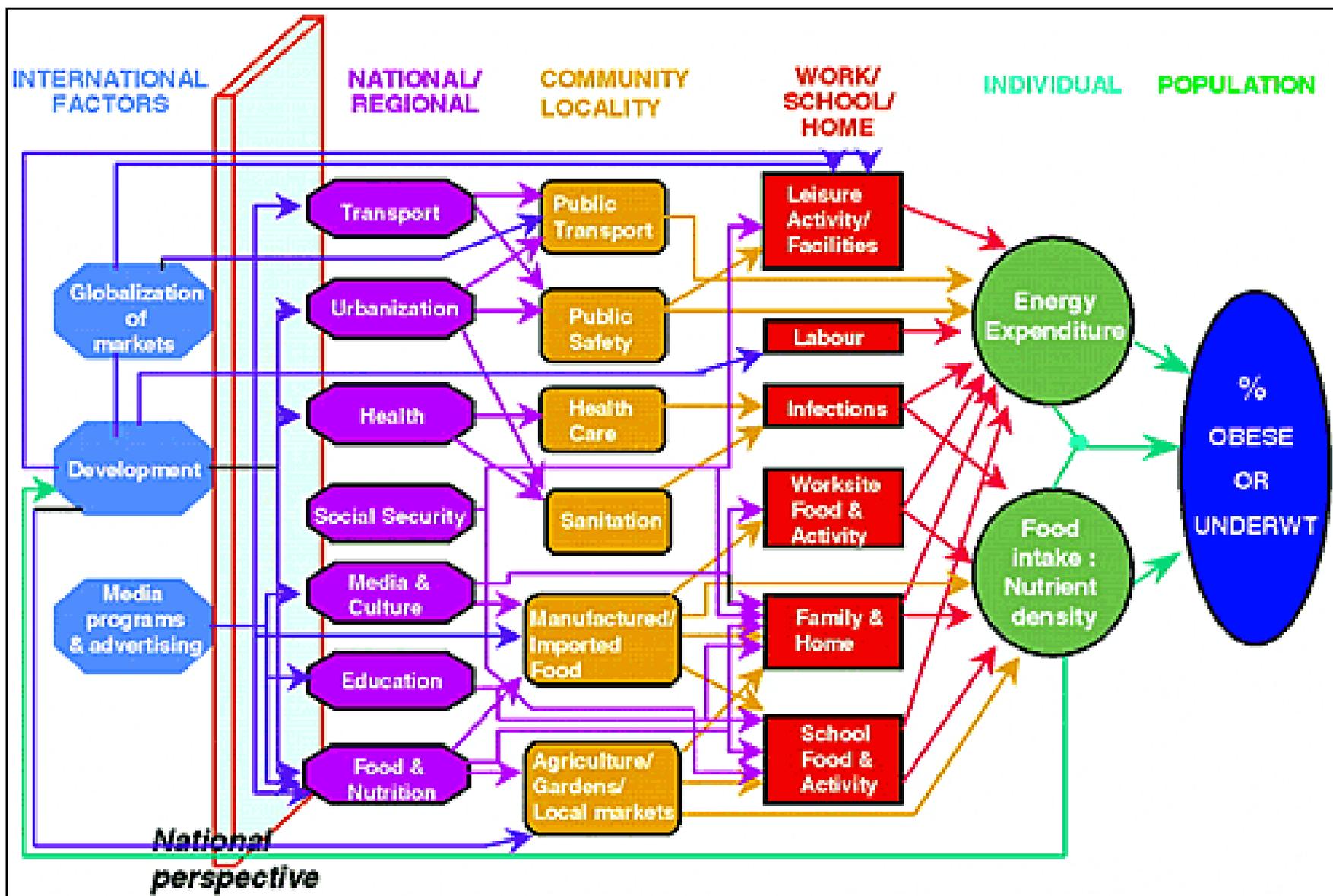
  - They may be transmitted to hosts via vectors**

# Classic Epidemiologic Theory

- **Host factors:**
  - Immunity and immunologic response
  - Host behavior
- **Environmental factors:**
  - Physical environment (heat, cold, moisture)
  - Biologic environment (flora, fauna)
  - Social environment (economic, political, culture)

# Web of Causation for the Major Cardiovascular Diseases





# Etiology of a disease

18

- ✓ **The sum of all factors that contribute to the occurrence of a disease**
- ✓ **Agent factors +Host factors +Environmental factors = Etiology of a disease**
- ✓ **The factor which can be modified, interrupted or nullified is most important.**

# Factors for disease causation

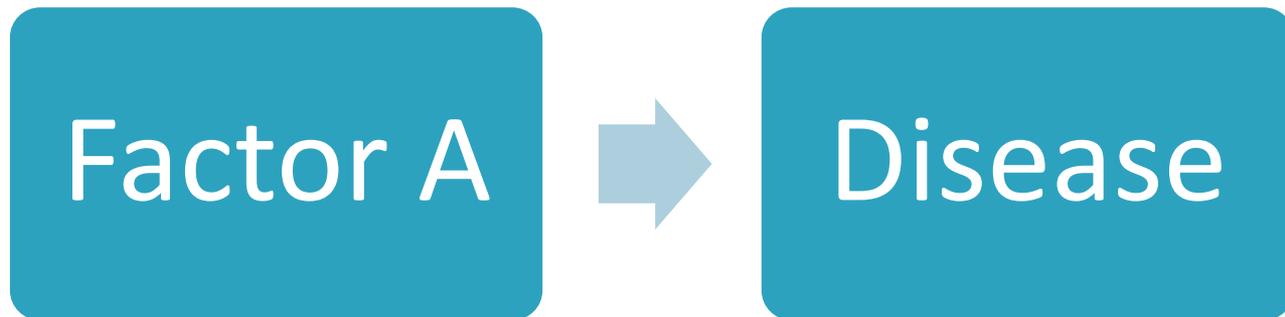
- **Sufficient factors:** one that inevitably produces disease (the presence of the factor always result in disease).
- **Necessary factors:** without which disease does not occur, but by itself, it is not sufficient to cause disease (the disease will not occur without the presence of the factor)

# Types of Causal Relationships

- **Four types possible:**
  - Necessary & sufficient
  - Necessary, but not sufficient
  - Sufficient, but not Necessary
  - Neither Sufficient nor Necessary

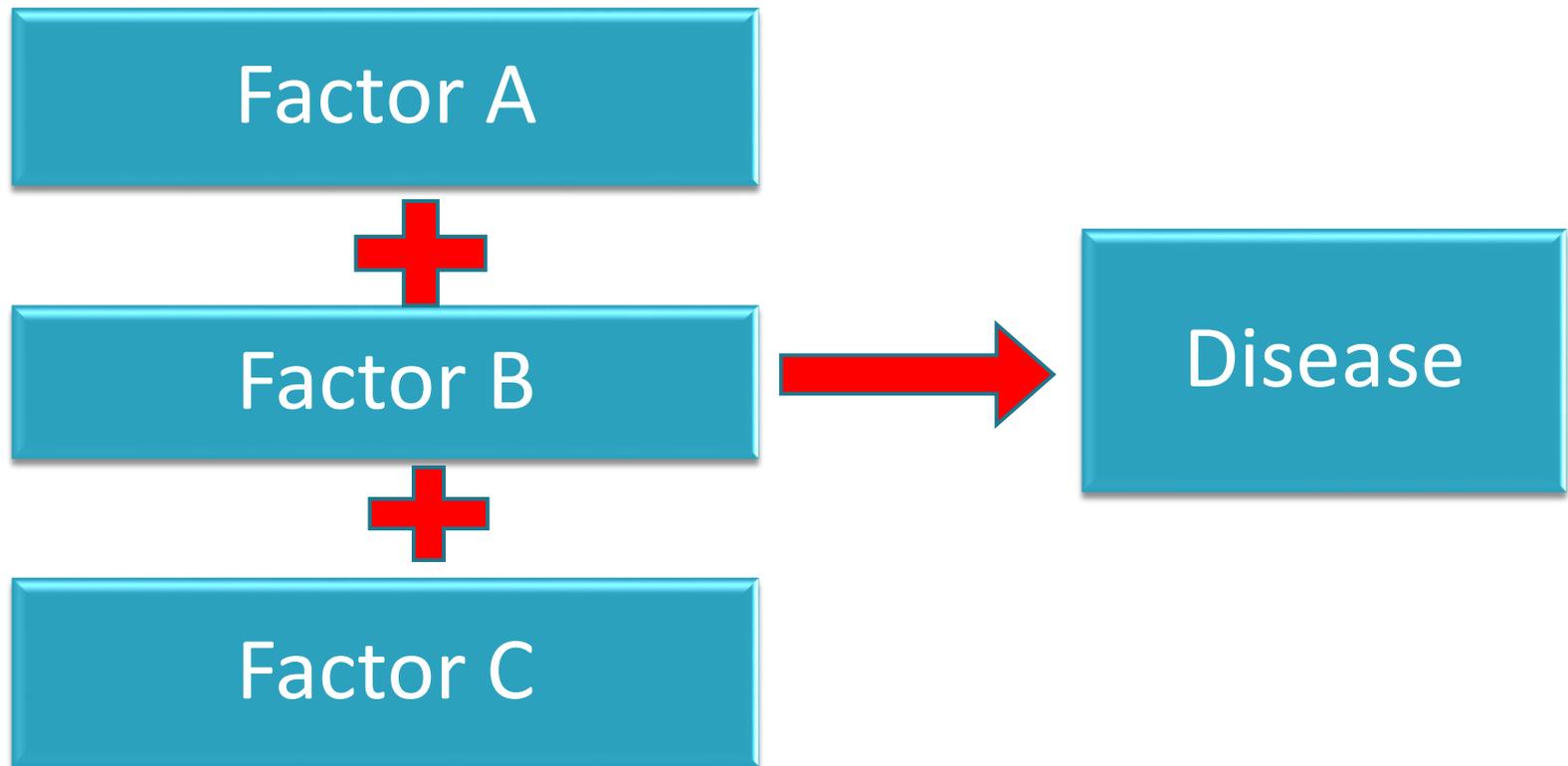
# I. Necessary & Sufficient

- Without that factor, the disease never develops (**factor is necessary**)
- and in presence of that factor, the disease always develops (**factor is sufficient**).
- Rare situation.



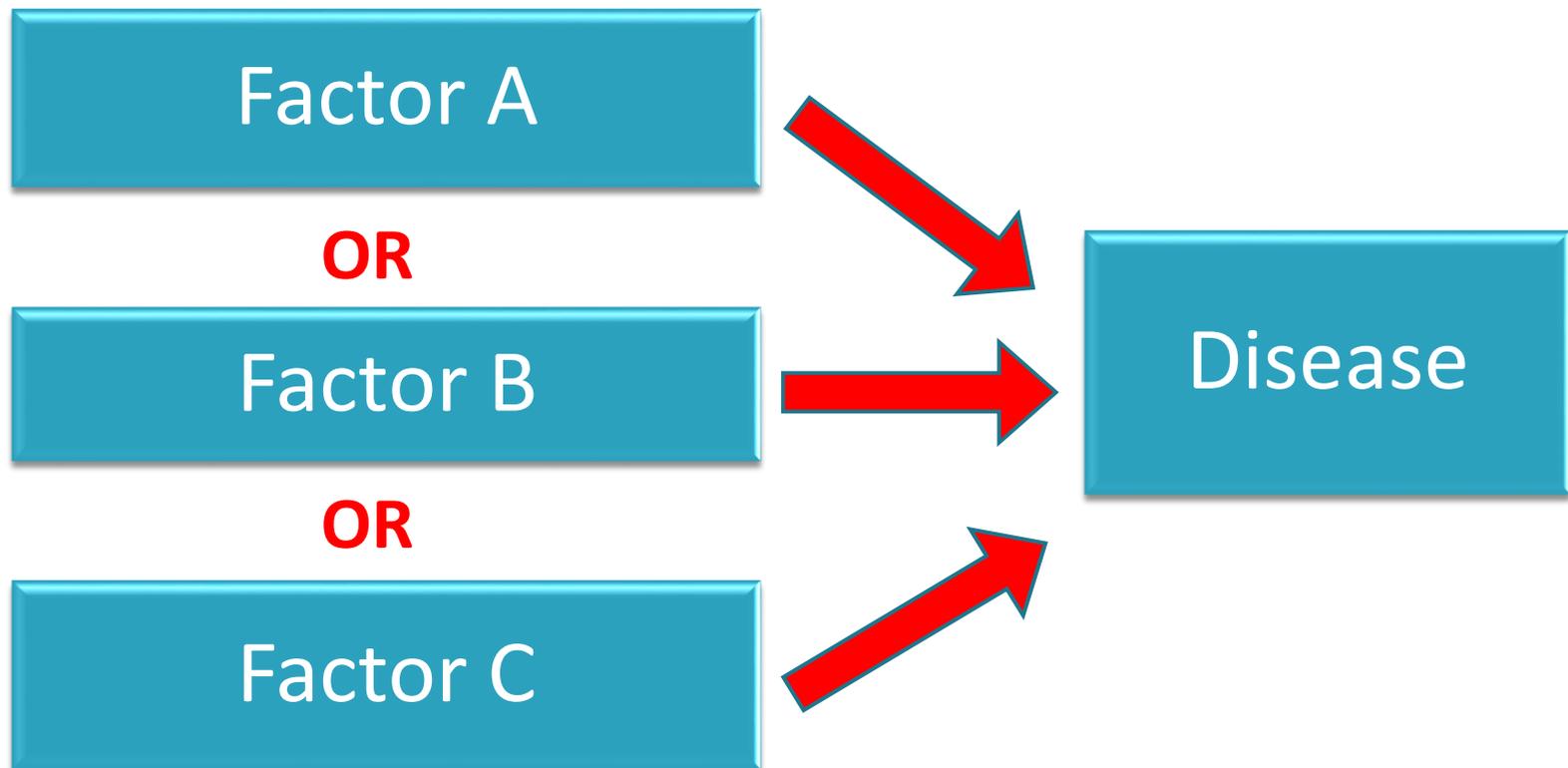
## II. Necessary, but not Sufficient

- Multiple factors are required, often in specific temporal sequence (cancer, initiator then promoter)
- Infectious diseases also.



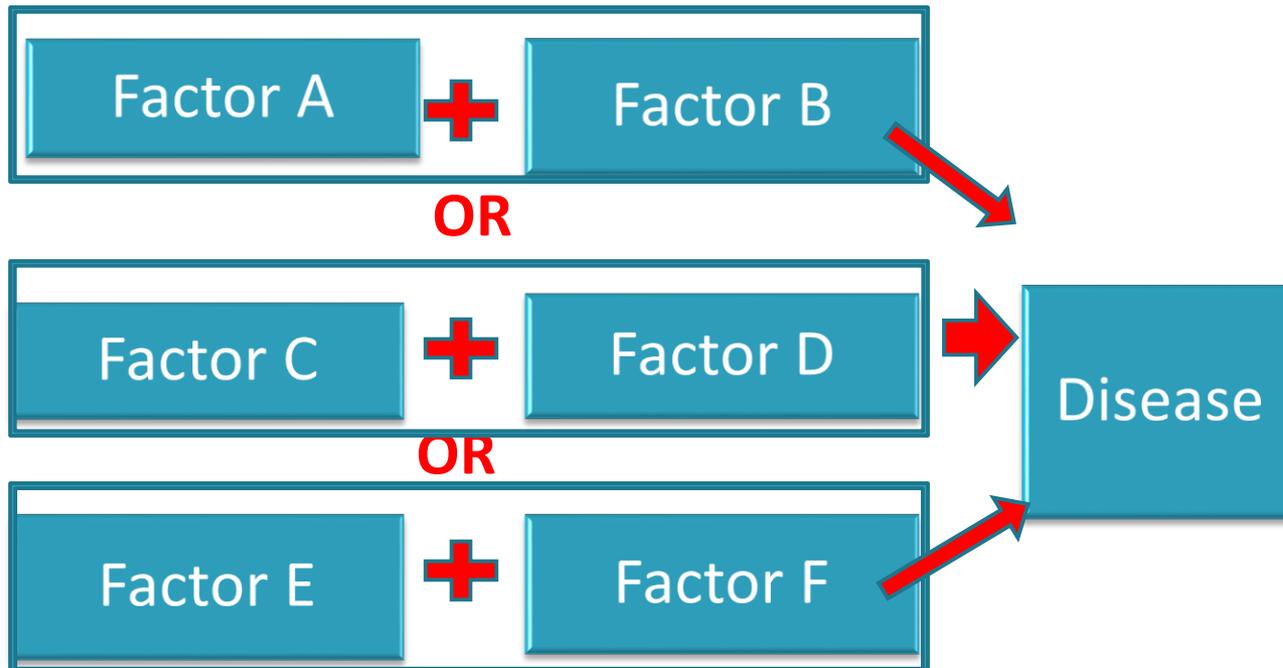
# III. Sufficient, but not Necessary

- Various factors independently can produce the disease (**Either radiation or benzene exposure can each produce leukemia without the presence of the other**).



# IV. Neither sufficient nor Necessary

- More complex model.
- Probably most accurately represents causal relationships that operate in most **chronic diseases**.



# Example....

- A researcher in his observational study found the presence of *Helicobacter pylori* in patients of duodenal ulcer!
  
- **Can we say that**
  - ▣ H.pylori causes duodenal ulcers?
- **Hypothesize that**
  - ▣ H.pylori **may** have a role in etiology of duodenal ulcers.
- For final proof there has to be a 'comparison'.
- Comparison would generate another summary measure which shows the extent of 'Association' or 'Effect' or 'risk'

# Process of establishing a “Cause & Effect” or “Exposure & Outcome” relationship

- Needs a research on the lines of ‘hypothesis testing’
- final establishment of an “exposure - outcome” relationship consists of a sequence of steps as follows :
- **Step 1:** ensure that the results of the study are accurate and not “spurious”.
  - ▣ Correct methods?
  - ▣ Validity, reliability preserved?
  - ▣ Bias?

# Process of establishing a “Cause & Effect” or “Exposure & Outcome” relationship

- **Step 2a:** do statistical results indicate association?-  
p value/ 95% CI.
- **Step 2b:** if not significant p value, may be b/c of  
low power of the study (smaller sample size)-  
The investigator should suggest additional studies  
using large sample (or else, a ‘meta - analysis’ type  
of study), rather than straightaway dismissing the  
‘exposure - outcome’ association as non - causal.

# Process of establishing a “Cause & Effect” or “Exposure & Outcome” relationship

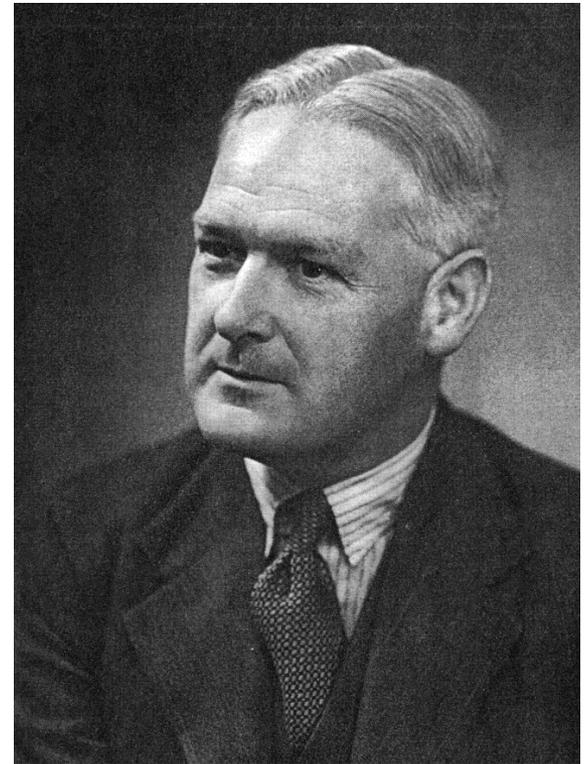
- **Step 3:** if statistically significant –evaluate as to whether this relationship is due to ‘indirect relationship’ with a third variable (confounder).

# Process of establishing a “Cause & Effect” or “Exposure & Outcome” relationship

- **Step 4:** if confounder excluded- now test this postulated “causal” relationship on the following **criteria of “causal association”**

# Sir Austin Bradford Hill, 1965

- *In what circumstances can we pass from [an] observed association to a verdict of causation? Upon what basis should we proceed to do so?*



# Guidelines for judging whether an association is causal

## Sir Austin Bradford Hill criteria

### □ Most Important criteria

1. **Temporality:** cause precedes effect
2. **Strength of association:** large relative risk
3. **Consistency:** repeatedly observed by different persons, in different places, circumstances, and times

# Guidelines for judging whether an association is causal

## □ **Additional supportive criteria**

4. **Biological gradient (dose response):** larger exposures to cause associated with higher rates of disease. And reduction in exposure is followed by lower rates of disease (reversibility).
5. **Biological plausibility:** makes sense, according to biologic knowledge of the time.
6. **Experimental evidence.**
7. **Other criteria: Analogy** (cause & effect relationship already established for a similar exposure or disease), **specificity** (one cause lead to one effect) and **coherence.**

# 1. Strength of association

**The larger the magnitude of association the more likely the exposure affects the risk of developing the disease.**

- ▣ Quantify how much the exposure increases the risk of disease.
- Epidemiologic Measures:
  - ▣ Risk ratio (RR), risk differences (AR)
- Example:
  - ▣ RR of lung cancer in smokers vs. non-smokers = 9
  - ▣ RR of lung cancer in heavy vs. light smokers = 20
  - ▣ Mortality from scrotal cancer among chimney sweeps compared to others = 200

## 2. Consistency

- **Definition:** The association is observed repeatedly in different persons, places, times, and circumstances.
- **Why Important?** If association is observed under different circumstances, with different samples and study designs, the more likely it is to be causal.
  - **Smoking associated with lung cancer in 29 retrospective and 7 prospective studies (Hill, 1965)**

# 3. Temporality

- Definition: The factor that is hypothesized to cause the disease must precede it in time.
- Why important?: A factor can co-occur with a disease and not cause it. In some cases, a factor might actually result from a disease.
- Epidemiology: Study design: Prospective cohort studies designed so that we know the exposure precedes the outcome.

# 4. Experiment

- Definition: Investigator-initiated intervention that tests whether modifying the exposure through prevention, treatment, or removal, results in less disease.
- Why Important?: Most epidemiologic studies are observational.
- RE. Epidemiology: Randomized clinical trials are closest to experiments in epidemiology.

# 5. Specificity

---

- Definition: The extent to which one exposure is associated with one outcome or disease.
- Why important?: Be certain that you identify the particular agent, or cause, that results in a particular outcome.

# 5. Specificity

- A single factor can cause several diseases (e.g., smoking associated with increased risk of lung cancer, small birth weight babies, etc.).
- Also, a single disease can be caused by many factors (e.g., heart disease).
- Bradford-Hill: Specificity should be used as evidence in favor of causality, not as refutation against it.
- **Example:**
  - Smoking associated with lung cancer, as well as other conditions (lack of specificity)
  - Lung cancer results from smoking, as well as other exposures.

# 6. Biological Gradient

- Definition: A “Dose Response” association. Persons who are exposed to greater amounts of a risk factor show increasingly higher “rates” of disease.
- A dose-response relationship provides support for causality, but the lack of this relationship does not mean lack of causality.
- **Example:**
  - Lung cancer death rates rise with the number of cigarettes/day smoked.
  - The 16 year risk of colon cancer was similar among women in each of the 5 levels of dietary fiber intake, from lowest to highest (Fuchs et al.,1999).

# 7. Biological Plausibility

- Definition: Knowledge of biological (or social) model or mechanism that explains the cause-effect association.
- Epidemiologic studies often identify cause-effect relationships before a biological mechanism is identified
  - E.g. In the mid 19<sup>th</sup> century when a clinician recommended hand washing by medical students & teachers before attending obstetric units, his recommendations were dismissed by medical fraternity as “doesn’t stand to reasoning”
  - E.g., John Snow and cholera; thalidomide and limb reduction defects).

**Bradford-Hill noted that biological plausibility cannot be “demanded”.**

# 8. Coherence



- Coherence - On the other hand, the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease.

# 9. Analogy



- Definition: Has a similar cause-effect association been observed with another exposure and/or disease?
- Why Important?: Important for generating hypotheses for the cause of newly-observed syndromes.

# From Association to Causation

