My reservoir is like the distance between Aden and Amman in al-Balqa

Al-Balqa Applied University

Faculty of Medicine

Epidemiology and Biostatistics

الوبائيات والإحصاء الحيوي (31505204)

Lecture 17+18

Disease causation and association

By

Tariq N. Al-Shatanawi

MPH PhD

3+4-8-2021

Epidemiological Principles

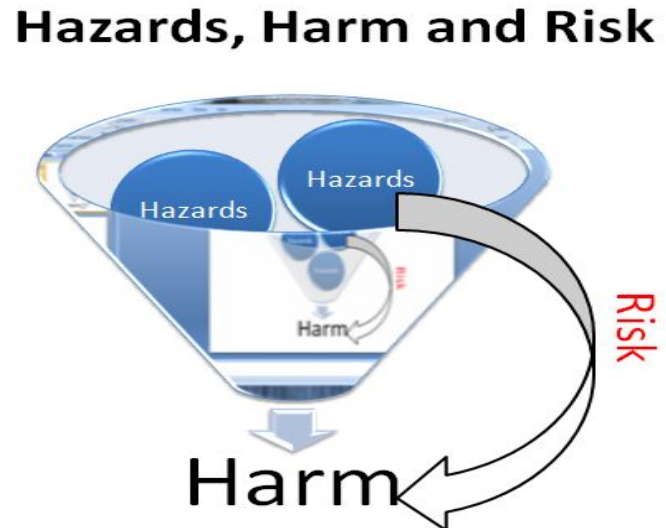
- **Diseases (or other health events) don't occur at random.**

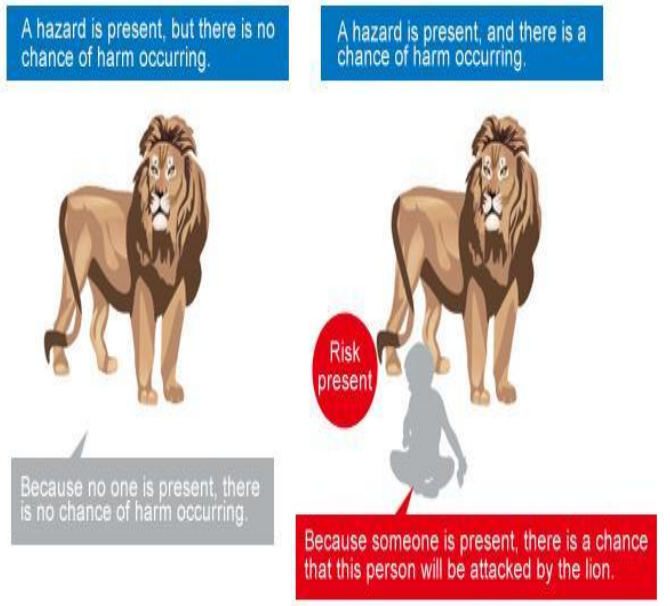
- **Diseases (or other health events) have causal and preventive factors which can be identified.**

- **Hazard** is something with the potential to cause harm, such as a substance, a piece of equipment, a form of energy, a way of working or a feature of the environment.

- **Harm** includes death and major injury and any form of physical or mental ill health.

- **Risk** it is a measure of the probability that the hazard (defined previously) will manifest some degree of harm.





Risk

- **Risk is the likelihood of an event occurring.**

Types of risk

- **Absolute: Incidence** of disease in any population.
- **Relative: Ratio** of the **incidence rate** in the group **exposed** to the hazard to the **incidence rate** in the **non-exposed** group.
- **Attributable: Difference in incidence rates between exposed and non-exposed groups.**

What is causality

IF-THEN rule that links two events.

Cause **Effect**

- **Causality (causation / cause-effect relationship):** relating causes to the effects they produce.

A cause?

- **The first and difficult question is, what is a cause?**
- **A cause is something which has an effect.**
- **In epidemiology a cause can be considered to be *something that alters the frequency of disease, health status or associated factors in a population.***

- **Cause:** An event, condition, characteristic (or a combination) which **plays an important role / regular / predicable change in occurrence of the outcome** (e.g. smoking and lung cancer).

- Causes may be **"genetic" and** / or "environmental" (e.g. many NCDs including: diabetes, cancers, COPD, etc).

Cause and effect

- **Cause and effect understanding is the *highest form of achievement of scientific knowledge*.**

- **Causal knowledge can help predict the outcome of an intervention and help treat disease.**

- **Quote Hippocrates *"To know the causes of a disease and to understand the use of the various methods by which the disease may be prevented amounts to the same thing as being able to cure the disease"*.**

Association and Causation

- **Association (relationship)**: Statistical dependence between two or more events, characteristics or other variables. **Positive association** implies a direct relationship, while **negative association** implies an inverse one. The presence of a statistical association alone does not necessarily imply a causal relationship.

- Association is defined as **" Occurrence of two variables more frequently than expected".**

- Association *between suspected cause and effect* ***does not*** *always mean a causal association.*

Types of Association

- **Association may be grouped into following three types;**

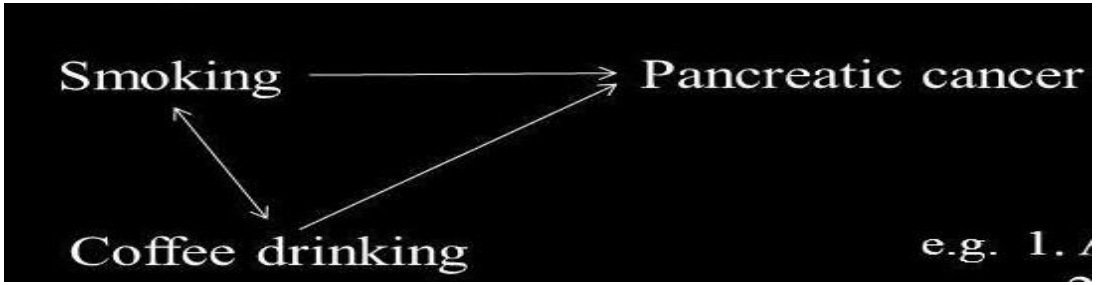
1. **Spurious Association** : When the observed association between suspected cause and effect **may not be real**.

   - **Example**- Perinatal mortality being high in hospital deliveries than home deliveries implying hospital is unsafe. The cause of spurious association is poor control of **Biases** in study.

Types of Association

2. **Indirect Association :** It is a ***statistical association*** between a factor of interest and a disease due to presence of another factor known **as Confounding Factor.**

- **Example :** Neonatal mortality (A) was found to be associated with maternal age above 30 years (B) and with birth order 4 and above (C).



Types of Association

3. Direct Causal Association : One to one and multifactorial.

- The association between the two attributes is not through the third attributes.

A. One to one causal Association :

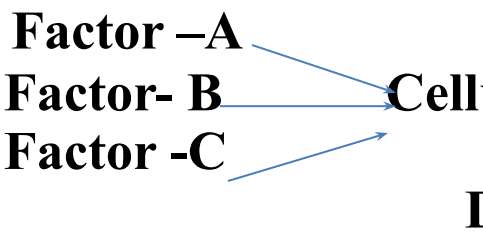
- Two variables /factors are considered to be causally related if any change in one is followed by a change in the other.

B. Multifactorial Causation :

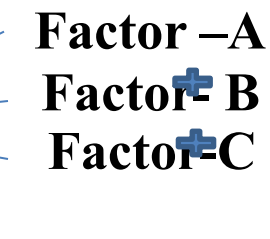
- Single causation theory **does not explain** causation of Non-Communicable diseases where multiple factors are involved in causation of disease viz; CHD and Ca-Lung.

- Two models presented below may explain multifactorial causation mechanism ;

1.



2.



Direct Vs indirect cause

Direct

Delta F508 mutation
DNA Polymorphism

Cystic Fibrosis

Indirect

High cholesterol

Artery thickening

Hemostatic factors

Myocardial infarction

Direct (Causal) association

a) One-to-One causal association

Two variables(**AB**) are stated to be causally related if a change in **A** is followed by a change in **B**.

b) Multifactorial association

Considered when the etiology is multifactorial.

All the causal factors can act individually or cumulatively to produce the outcome.

A Unifying Model of Causal Relationships

❑ The 2 Components:

- **Sufficient Cause**
  - precedes the disease.
  - if the cause is present, the disease always occurs.

- **Necessary Cause**
  - precedes the disease.
  - if the cause is absent, the disease cannot occur.

Necessary Condition

- ***Must be there for the effect to be true.***
- ***If absent, cannot occur.***
- ***No oxygen, no combustion.***
- ***No seeds, no plants to grow.***
- ***Car runs only if gas in tank.***

Sufficient Condition

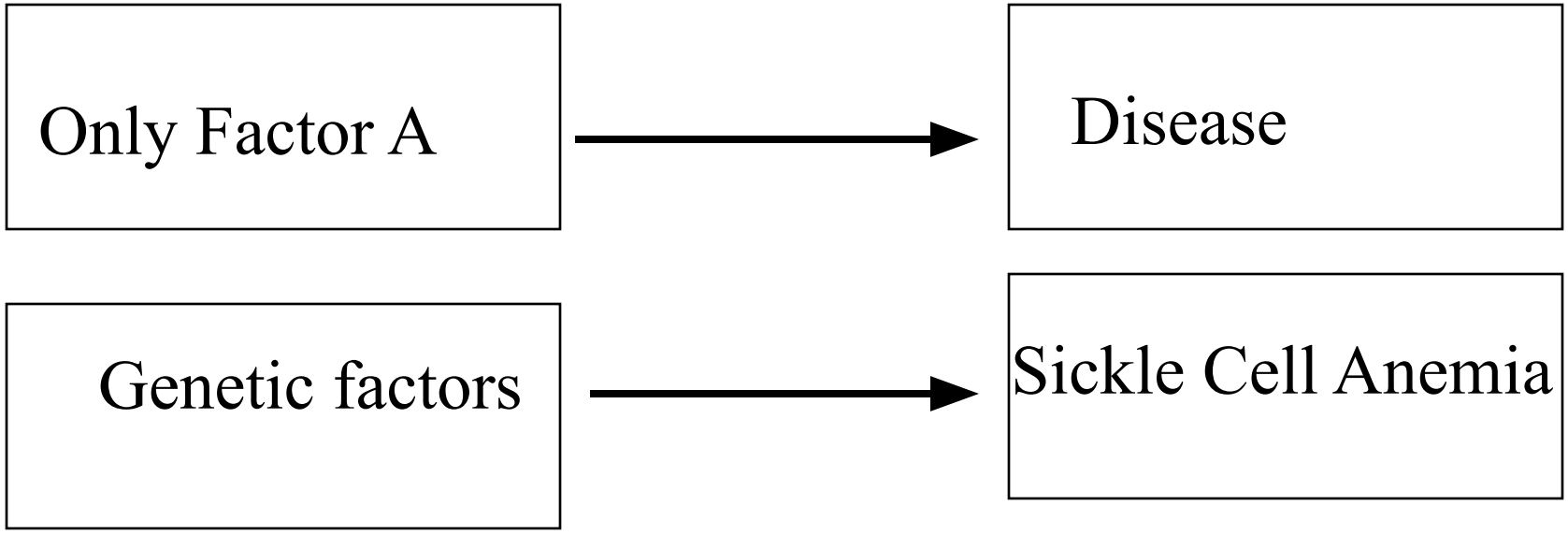
- **If factor A is present, Factor B is absent / Factor A is sufficient for death.**

- Concept of Necessary vs. Sufficient Causes provides a theoretical framework for causation of all disease.

The 4 Models of Causal Relationships

1. Necessary and Sufficient*

- **If the Factor is Present, the Disease will Always Occur.**
- **Without the factor, the Disease NEVER develops.**
- **Most infectious diseases will not cause illness in everyone, and not all heavy smokers develop Lung Ca.**



* RARELY OCCUR

Necessary & Sufficient

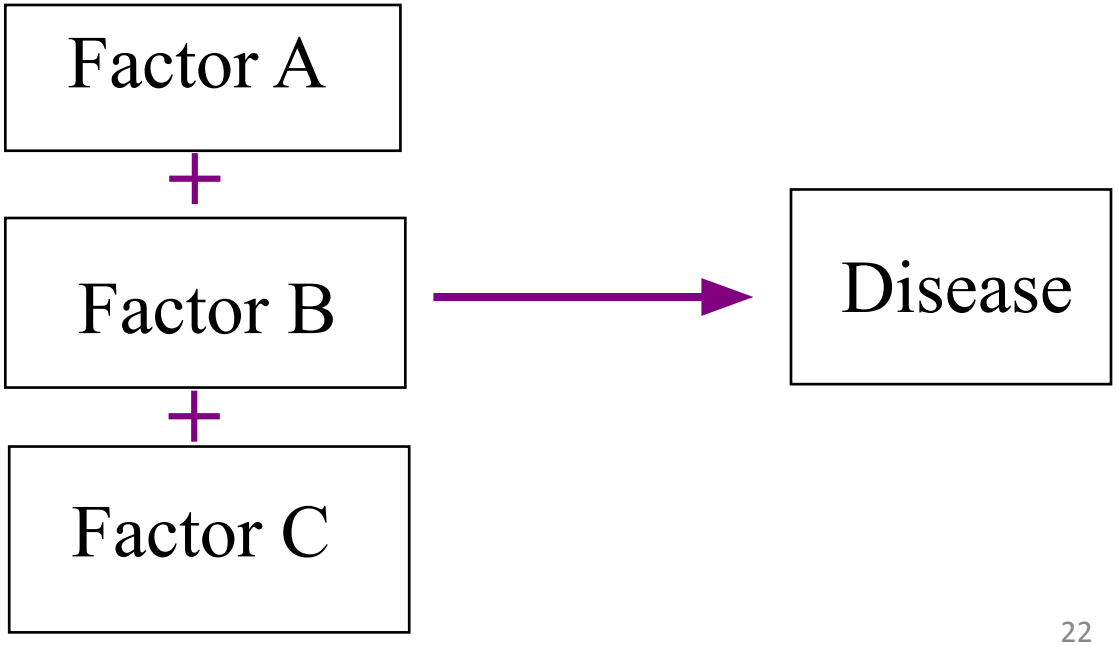
- **Something can serve as both necessary and sufficient- "You will get Malaria if and only if you are bitten by a mosquito carrying the germ. Malaria ≈ Mosquito (germ).**

- **If you have Malaria you must have been bitten by the mosquito with the germ.**

2. Necessary but Not Sufficient

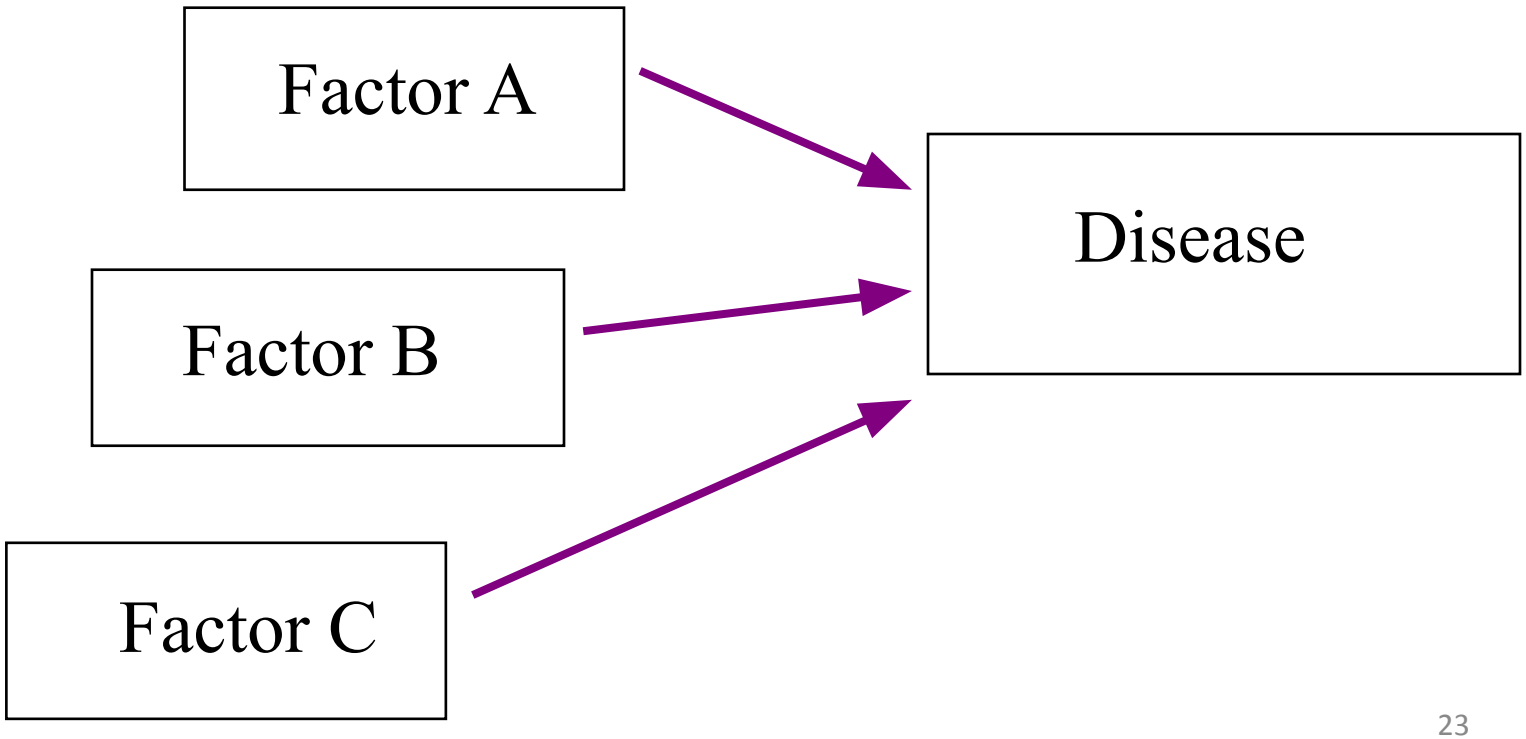
- **Each factor necessary but not in itself sufficient to cause the illness in itself.**
- **All are necessary to cause disease, but individually, none are sufficient to cause the disease.**

- **Each Risk factor alone Cannot Cause Disease.**



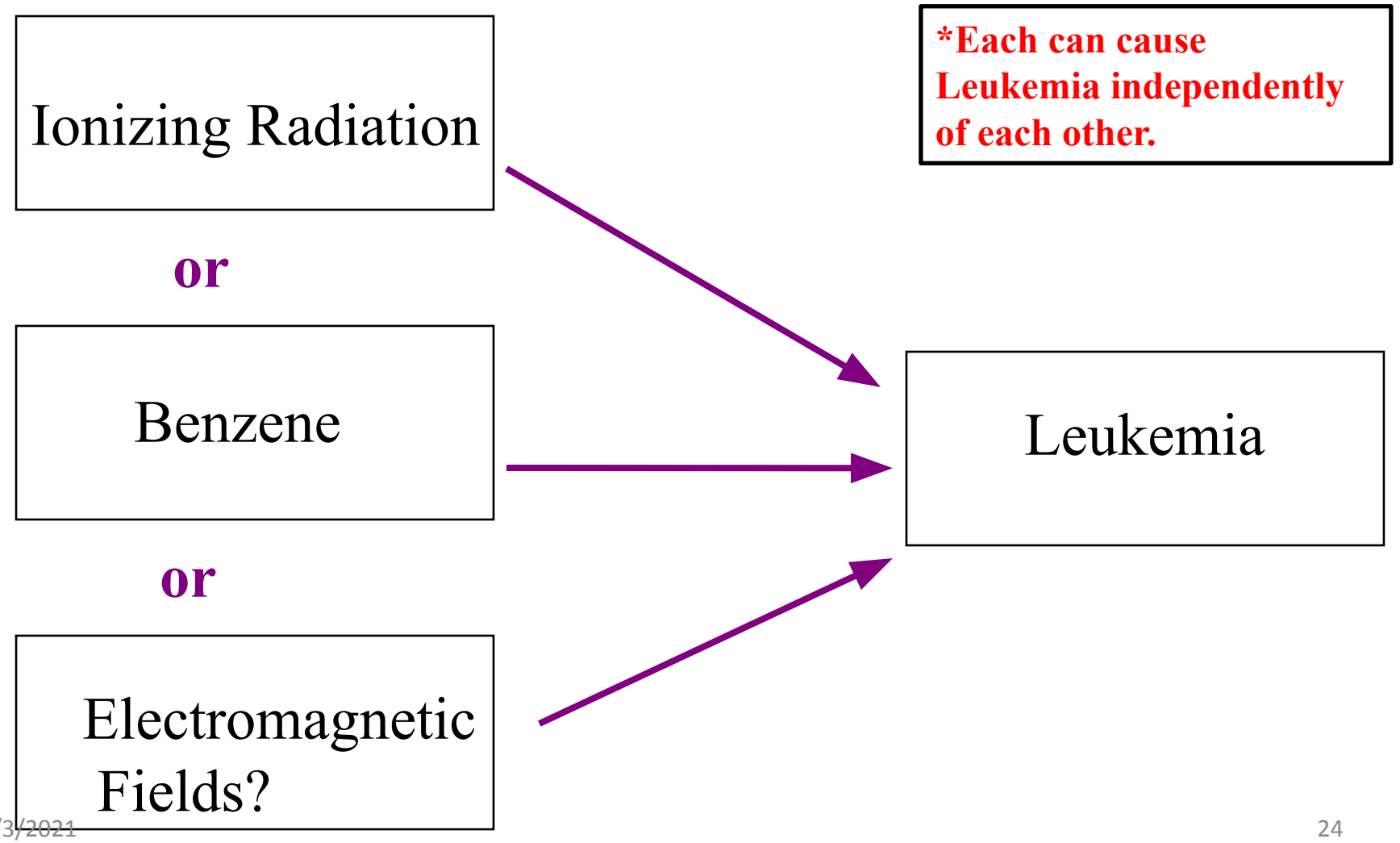
3. Sufficient but Not Necessary

- **Each risk factor Sufficient to cause disease, but not all risk factors necessary in disease causation (If Factor C is absent, disease can still occur).**

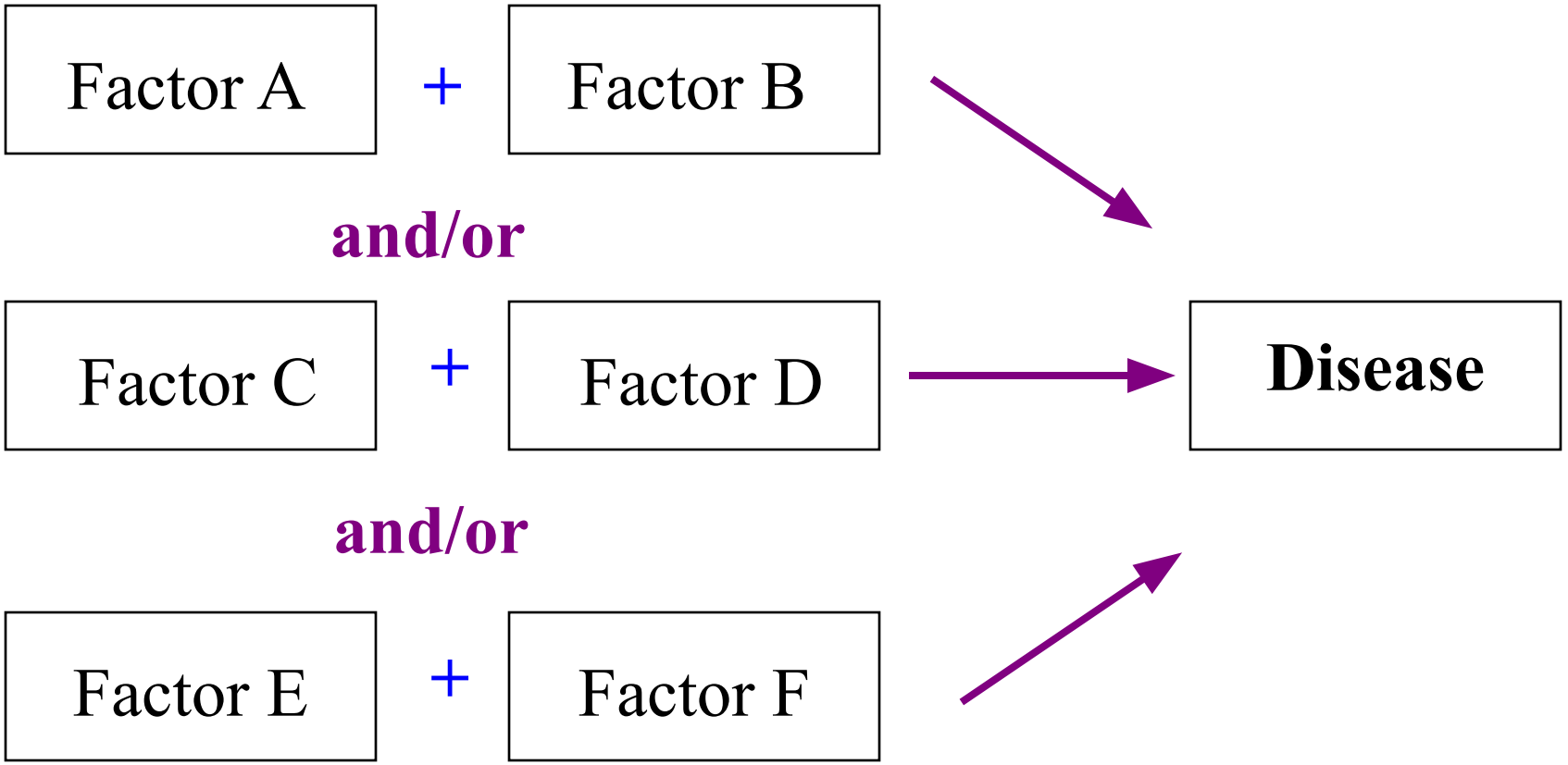


3. Sufficient but Not Necessary - Example

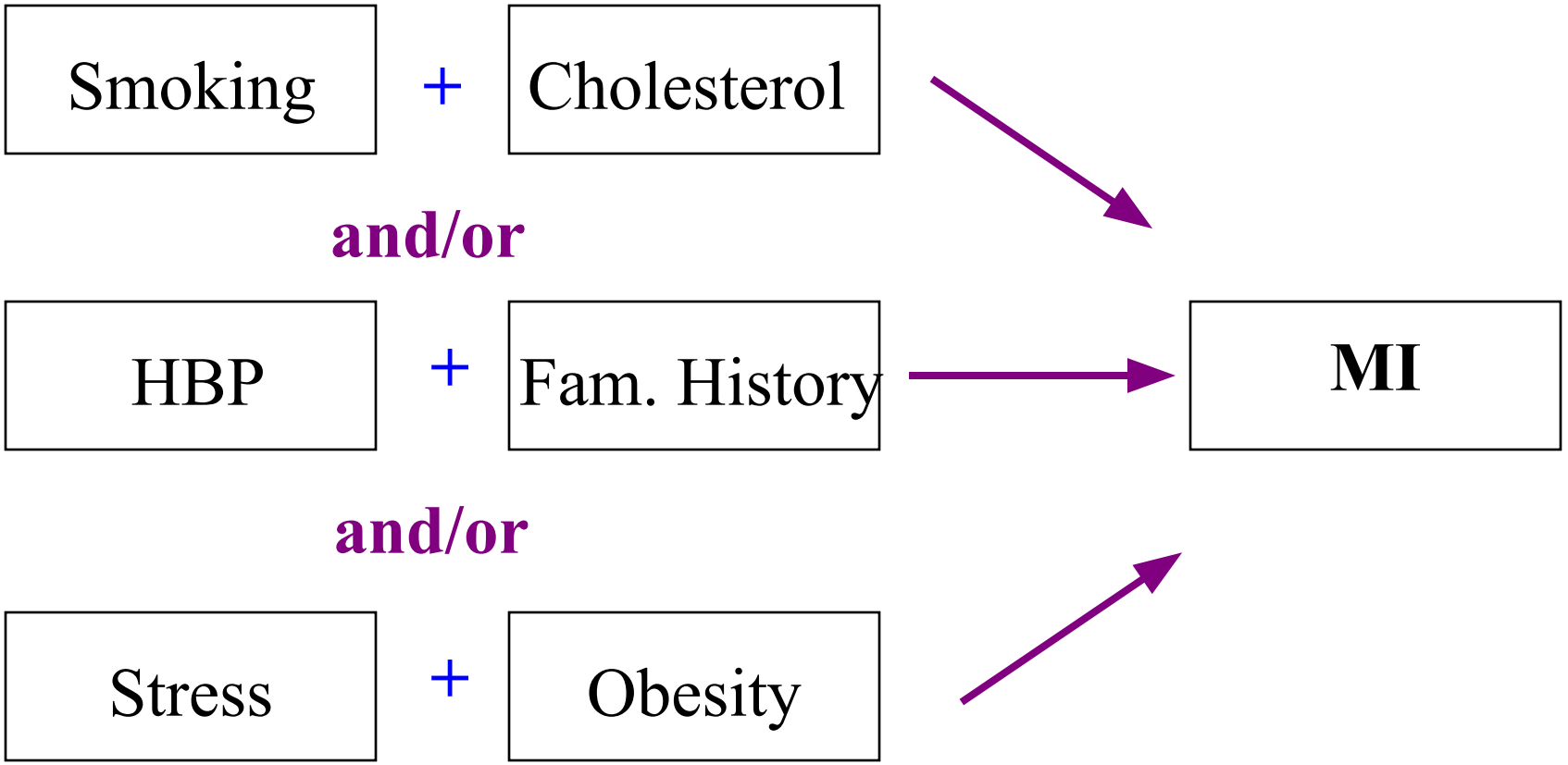
- **If exposed to enough Radiation, other 2 Risk factors not necessary to cause the illness.**



4. Neither Sufficient Nor Necessary



4. Neither Sufficient Nor Necessary - Example



Hill's criteria

1. Strength of association
2. Temporal relationship
3. Distribution of the disease
4. Gradient
5. Consistency
6. Specificity
7. Biological plausibility
8. Experimental models
9. Preventive trials

Criteria for Assessing Causation

Temporal relationship

- Exposure precedes the disease (**Exposure to Risk factor occurred before Illness onset**).

Strength of the Association

- Measured by the Relative Risk (either the Rate Ratio or the Odds Ratio) (**Higher the relative Risk, more likely Causal**).

Dose-response Relationship

- As the dose of exposure increases the risk of disease also increases. (**strong evidence for Causal relationship).**
- Example: Cigarette Smoking and Lung Ca.

Criteria for Assessing Causation

Biologic plausibility

- Does the association fit with what we know about the underlying biology.

- Sometimes we know little or nothing about the underlying biology ("Black Box" epidemiology). **Example** – Asbestosis and Lung Ca.

- Causal association is substantiated **if biological plausibility is present.**

- The positive association of **intestine, colon and rectum with food intake is biologically plausible**.

- The **cigarette smoking** and **lung cancer** is also biologically plausible as it is well established that tobacco contains carcinogen which cause Ca-lung.

Consistency of Association (Replication of the Findings)

- The **result from single study is seldom sufficient to establish causal association.**
- Results replicated in other studies. **(The relationship is verified by repeated studies).**
- An association has to **be replicated and confirmed by different investigators**, in different populations using different design in order to be established as causal association.

- More than fifty retrospective studies and at least nine prospective studies in different countries have shown a consistent association between cigarette smoking and subsequent development of lung cancer has lent support to a causal association.

Relating

- **Exposures: causes, risk factors, independent variables to…**
- **Outcomes: effects, diseases, injuries, disabilities, deaths, dependent variables.**
- **Statistical association versus biological causation: cause-effect relationship.**

Summary

- **Cause and effect understanding is the highest form of scientific knowledge.**

- **An association between disease and the postulated causal factors lies at the core of epidemiology.**

- **Demonstrating causality is difficult because of the complexity and long natural history of many human diseases and because of ethical restraints on human experimentation.**

Summary....

- **All judgements of cause and effect are *tentative.***
- **Be alert for error, the play of chance and bias.**
- **Apply criteria for causality as an aid to thinking.**
- **Look for corroboration of causality from other scientific frameworks.**

Definitions

- **Deterministic causality:** cause closely related to effect, as in "necessary" / "sufficient" causes.

- **Component causes**: Together they constitute a sufficient cause for the outcome in question.

  In CDs, this may include the biological agent as well as environmental conditions (e.g. TB, measles).

  In NCDs, this may include a whole range of genetic, environmental as well as personal / psychosocial / behavioral characteristics (e.g. diabetes, cancers, IHD).

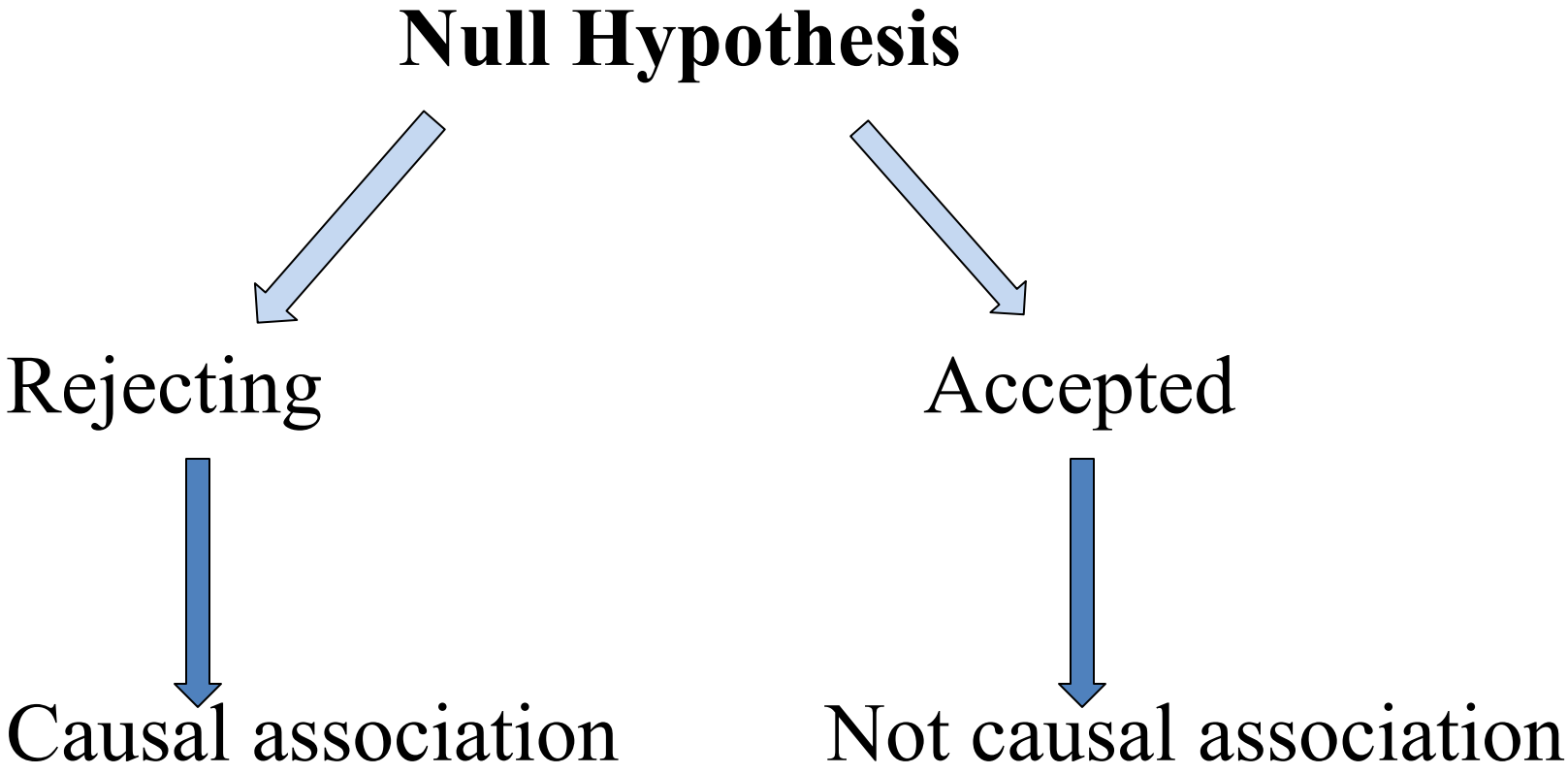
Expressions of Strength of Association

- **Quantitatively:**
  - **Effect measure (OR, RR): away from unity (the higher, the stronger the association).**
  - **P-value (at 95% confidence level): less than 0.05 (the smaller, the stronger the association).**
- **Qualitatively:**
  - **Accept alternative hypothesis: an association between the studied exposure and outcome exists.**
  - **Reject null hypothesis: no association exists.**

Analytical Methods

- **Measures of association /strength of association.**
- **Testing hypothesis of association.**
- **Controlling confounders.**

Testing hypothesis of association



Measures of association / strength of association

- Measure the strength of association between the exposure and outcome, e.g. How likely are cigarette smokers likely to develop lung cancer?

- **Ratio measures**
  - **Relative risk**
  - **Odds ratio**

- **Difference measures**
  - **Attributable risk**
  - **Population Attributable risk**

Strength of association

- **Relative risks/Odds ratio**

Risk ratio	Interpretation
< 1	Protective
0.9-1.1	No association
1.2- 1.6	Weak Causal association
1.7- 2.5	moderate causal association
>2.6	Strong causal association

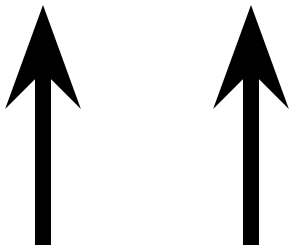
Exploring Linear Relationships

- Researchers use **crosstabs and comparison of means between two variables to see if there is a relationship.**

- If we see **some differences that suggest there is a relationship, the next steps is to determine how strong it is.**

Direction of Relationship Revisited

- **Plus sign: direct relationship**
  - Both variables change in the same direction.
  - **Example:** as **driving speed increases, death rate goes up.**



Direction of Relationship Revisited

- **Minus sign: inverse relationship**
  - Both variable change but in the **opposite direction.**
  - **Example**: as **age increases, health status decreases**

Measures of Association

- **How strong is the association?**

- Several different measures of association
  - **Some measures of association range from 0 to 1**
  - **Others range from minus1 to plus 1**

How To Interpret Measures of Association

- Measures of Association get interpreted in a similar way:

 Perfect Relationship = 1
  - Closer to 1: strong relationship
  - .5 moderate/strong

 Closer to 0: no relationship
  - .2 some/slight

How To Interpret Measures of Association

- Interpreting measures of association that have a minus sign:
  - Minus sign indicates an inverse relationship (meaning as one variable goes up, the other goes down)
    - As age increases, memory decreases
  - For example, -.9 is a very strong relationship (almost perfect relationship because it is close to -1), but it is an inverse relationship because it has a minus sign.

Epidemiologic Measures of Association

- **Compute & Interpret Relative risk (RR) & Odds ratio (OR) as a measure of association between exposure and Disease.**

- **Association: A statistical relationship between two or more variables.**

- **A relative risk of:**
  - **1.0 (or close to 1.0)** means the risk of disease is similar in the exposed and unexposed group and exposure is not associated with disease.
  - **Greater than 1.0** means the risk of disease is greater in the exposed than the unexposed group and the exposure could be a **risk factor** for the disease.
  - **Less than 1.0** means the risk of disease is less in the exposed group than the unexposed group and the exposure could be a **protective factor.**

Association between exposure & Disease

- **Objective:**

  - **To determine whether certain exposure is associated with a given disease.**

- **Methodology:**

  - **Use one of the epidemiologic study designs**
    - **Cohort**
    - **Case-control**

OR in case-control and cohort studies

- **Cohort study**

Ratio of the proportion of exposed subjects who developed the disease to the proportion of non-exposed subjects who developed the disease.

- **Case-control study**

Ratio of the proportion of cases who were exposed to the proportion of controls who were non-exposed.

Cohort Study

- **Assess the cumulative incidence (CI_{E+}) of disease in an exposed group (absolute Risk).**
- **Assess the cumulative incidence (CI_{E-}) of disease in unexposed group (absolute Risk).**

❑ ***e.g. Coronary Heart Disease (CHD) Risk among Smokers***

- **1-year risk of CHD among smokers (CI_{E+})***

	CHD		
	Yes	No	Total
Smokers	84	2916	3000

CI_{E+} = 84/3000 = 28/1000/yr (1-risk of CHD among smokers)

Cont.

CHD Risk among non-smokers

- **1-year risk of CHD among non-smokers (CI_{E-})**

	CHD		
	Yes	No	Total
• Non-smokers	87	4913	5000

- **CI_{E-} = 87/5000=17.4/1000/yr (1-yr risk of CHD among non-smokers).**

Cont.

Assessment of Excess Risk (Two methods)

a. ***Ratio***

RR (Ratio of two risks; Risk Ratio; Relative Risk)

$$CI_{E+} / CI_{E-} = 28/17.4 = 1.6$$

Interpretation of RR

Smokers were 1.6 times as likely to develop CHD as were non-smokers or 60% increase in risk of CHD in smokers vs non smokers

b. ***Difference***

Difference of two risks (Risk Difference)

$$CI_{E+} - CI_{E-} = 28.0 - 17.4 = 10.6$$

Derivation of OR in Cohort study

- $P_{D^+|E^+}$ = (exposed developed the disease) = **a/(a+b)**
- $P_{D^-|E^+}$ = (exposed did not develop the disease) = **b/(a+b)**

Odds of developing disease among exposed =

$$P_{D+|E+}/1-P_{D-|E+} = \underline{a/(a+b)} \qquad b/(a+b) = a/b$$

- $P_{D^+|E^-}$ = (non-exposed developed the disease) = **c/(c + d)**
- $P_{D^-|E^-}$ = (non-exposed did not develop the disease)= **d/(c + d)**

Odds of developing disease among non-exposed =

$$= P_{D+|E^-}/1-P_{D+|E^-} = \underline{c/(c+d)}$$
$$d/(c + d) = c/d$$

Odds ratio **a/b : c/d = ad/bc**

Diseases			
	(D)	(~D)	Total
Exposed	A	B	A+B
Non-exposed	C	D	C+D

The odds ratio

ad/bc

	Stroke (D)	No Stroke (~D)	
Smoker (E)	15	35	50
Non-smoker (~E)	8	42	50

$$OR = \frac{ad}{bc} = \frac{15*42}{35*8} = 2.25$$

Interpretation: there is a 2.25-fold higher odds of stroke in smokers vs. non-smokers.

Odds Ratio (OR)

Contingency (or 2 x 2) Table

	Cases	Controls	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$OR = (a/c) / (b/d)$$
$$= (a*d) / (b*c)$$

OR in case-control study

- In case-control study RR cannot be calculated directly to determine the association between exposure and disease.

- Don't know the risk of disease among exposed and un-exposed since we start recruiting cases and controls.

- Can use OR as measure of association between exposure and disease in a case control study.

OR in case-control Study

- Probability of case being exposed = P_{case}
- Probability of case being non-exposed =1-P_{case}
- Odds of case being exposed = P_{case}/1- P_{case}

- Probability of control being exposed = $P_{control}$
- Probability of case being non-exposed =1-$P_{control}$
- Odds of control being exposed = $P_{control}$/ 1-$P_{control}$

Derivation of OR in case-control Study

- **Probability of being exposed among cases = a /(a + c)**
- **Probability of being non-exposed among cases) = c /(a + c)**

❑ ***Odds of being exposed among cases*** **= a/c**

- **Probability of being exposed among controls = b/(b + d)**
- **Probability of being unexposed among controls = d/(b + d)**

❑ ***Odds of being exposed among controls*** **= b/d**

OR = ad/bc

Example
OR in case-control Study

- **Past surgery HCV status**

	HCV+	HCV-
• Yes	59	168
• No	54	48
»	113	216

- **Odds of Past surgery among HCV^+**

P_1 (Surgery among HCV^+) = 59/113

$1\text{-}P_1$ (No surgery among HCV^+) = 54/113

Odds of surgery among HCV^+) = 59/54 = 1.09

- **Odds of Past surgery among HCV^-**

P_2 (Surgery among HCV^-) = 168/216

$1\text{-}P_2$ (No surgery among HCV^-) = 48/216

Odds of surgery among HCV^- = 168/48 = 3.5

OR = 3.50/1.09 = 3.21

When is the OR a good estimate of RR?

- **In CCS, only OR can be calculated as measure of association.**

- **In Cohort study, either RR or OR is a valid measure of association.**