Association & Causation in epidemiological studies

DR. SIREEN ALKHALDI, DRPH
COMMUNITY MEDICINE 2023/ 2024
SCHOOL OF MEDICINE
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).

September 27, 2005 – New York Times - By GINA KOLATA

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:** (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 causal association**

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)!

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)

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? Bias?
- **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)-

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

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.

- **Step 3**: if statistically significant –evaluate as to whether this relationship is due to ‘indirect relationship’ with a third variable (confounder).
- **Step 4**: if confounder excluded- now test this postulated “causal” relationship on the following **criteria of “causal association”**

Bias and Confounding

If an association is observed, the first question asked must always be …

“Is it real?”

While the results of an epidemiological study may reflect the true effect of an exposure(s) on the development of the outcome under investigation, the findings may in fact be due to an alternative explanation.

Bias and Confounding

Such alternative explanations may be due to the effects of bias or confounding which may produce spurious results, leading us to conclude:

1. The existence of a valid statistical association when truly association does not exist.
2. The absence of an association when an association is truly present.

These factors need to be considered at both the design, conduct, and analysis stages of an epidemiological study so that their effects can be minimized as much as possible.

Bias

Bias is a systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease (Schlesselman and Stolley, 1982).

"Error" indicates that it is most probably unintentional.
"Systematic " implies that once it is introduced into the study, it cannot be fixed.

- ✓ The effect of bias will be an estimate either above or below the true value (>RR or <RR), depending on the direction of the systematic error. So, it affects the **validity** of the study (the degree to which the measurement reflects the true value in the population).

- ✓**Two types: Selection bias and information bias.**
- ✓**Can be avoided by defining criteria for selecting cases and controls, and exposed and non-exposed.**

Bias

Selection bias is a method of participant selection that distorts the exposure-outcome relationship from that present in the target population. Selection bias occurs when there is a systematic difference between either:

1. Those selected to participate in the study and those who do not OR
2. Those selected in the treatment group and those in the control group

Information bias results from systematic differences in the way data (information) on exposure or outcome are obtained from the various study groups (exposed vs non-exposed) (diseased vs non-diseased).

- **This yields systemic errors in the measurement of exposure or outcome. This will affect the nature of true association (recall bias).**

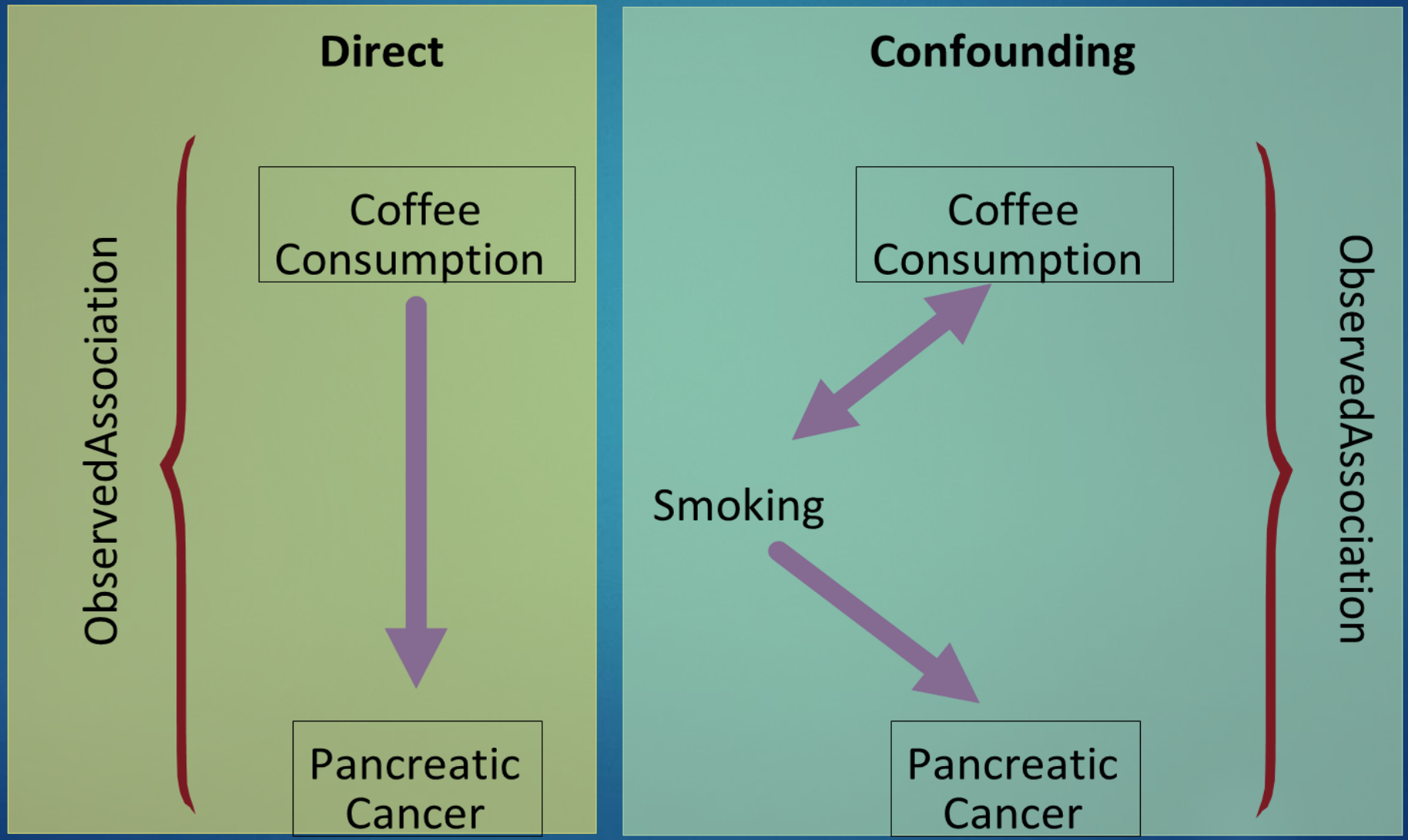
Confounding

Confounding occurs when the observed association between exposure and disease differs from the truth because of the influence of the third variable.

Confounder must be:

1. Risk factor for the disease independently
2. Associated with exposure under study
3. The variable should not lie on the causal pathway between exposure and disease.

Confounding



Confounding

Bias is a systematic error in a study and cannot be fixed if introduced into the study.

Confounding may lead to errors in the conclusion of a study, but, when confounding variables are known, the effect may be fixed (corrected, accounted for, controlled for).

Controlling of confounding at the design stage: restriction, matching and randomization.

Controlling at the analysis stage: stratification, multivariate analysis, and standardization.

Is there an association?

To conclude that an association between exposure and disease outcome exists:

- The study must have adequate sample size (power)
- The study must be free of bias
- The study must be adjusted for possible confounders

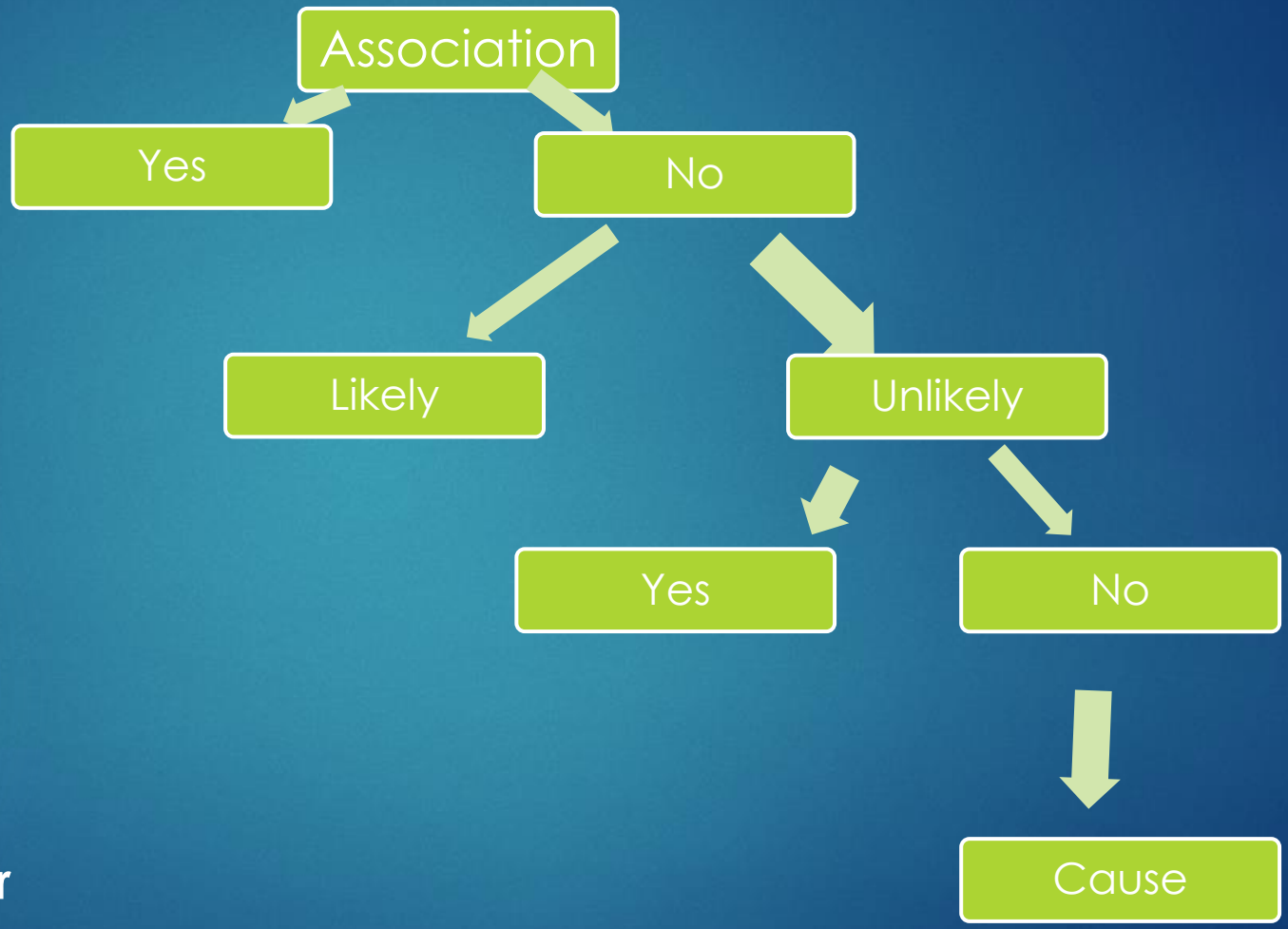
From Association to Causation

Bias in selection or measurement

Chance

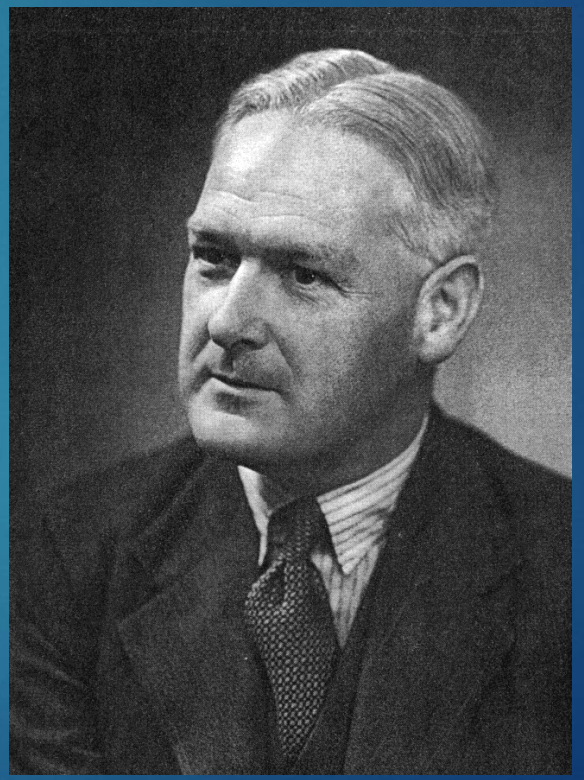
Confounding

Hill's criteria for Causality



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

Hill's Criteria: Nine criteria useful in establishing epidemiologic evidence of a causal relationship between a presumed cause and an effect:

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

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** (not seriously conflict with the generally known facts of the natural history and biology of the disease).

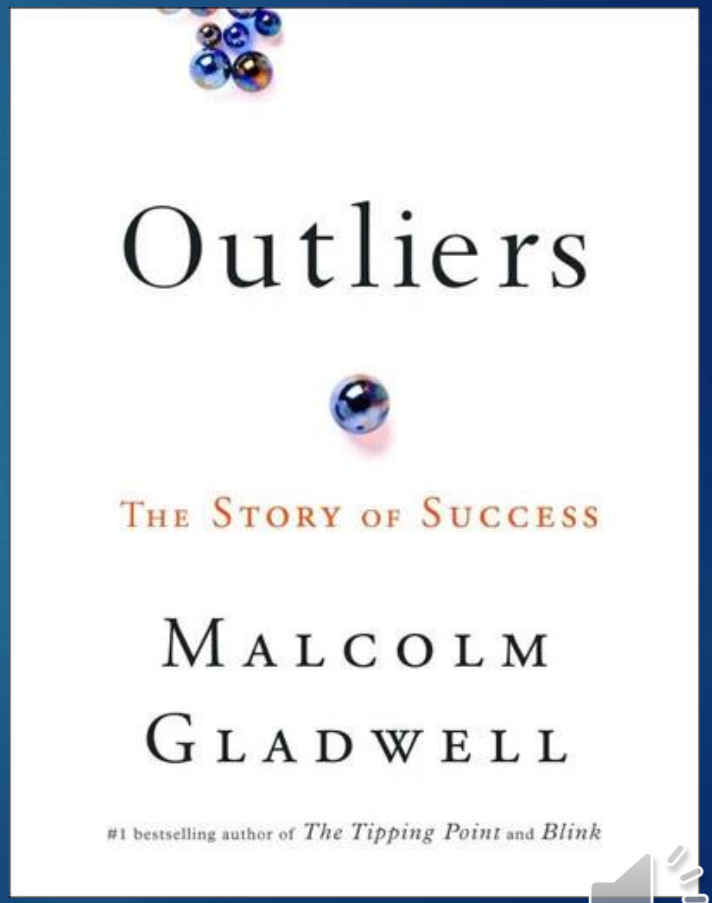
External Reading

Read the Introduction of the book "OUTLIERS, The Story of Success" for Malcolm Gladwell.

"The Roseto Mystery"

Your assignment is to find out why Dr. Wolf rarely found any one from Roseto village under Sixty- five with heart disease.

What was the protective factor??????



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 19th 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.