Modified slide

Modified by: Hamsa Taha
Corrected by: Ghada Al-zoubi

Association & Causation in epidemiological studies

DR. SIREEN ALKHALDI
COMMUNITY MEDICINE 2020/ 2021
SCHOOL OF MEDICINE
THE UNIVERSITY OF JORDAN

September 27, 2005 – New York Times -An article with a title, By GINA KOLATA:

''Which of these foods will stop cancer?''

well,we can't judge that so fast..[not so fast]

Cancer patients always ask what to eat to reduce their chances of dying from the disease.

Diet messages are everywhere:

- NCI(national cancer institute): 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 with epidemiological designs 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 do epidemiological studies,we** observe **associations after doing statistical analysis.**
- We infer (guess, speculate, reach to a conclusion) about **causes of certain disease according to what association we have found..**

ASSOCIATION

Definition: the concurrence of two variables together more often than would be expected by chance.

Types of Associations:

1. **Spurious Association:** **which is mistaken association, the study leads us for association which is not true** (for example:Shoe size and reading performance for elementary school children)-check next slide ^_^
2. **Indirect Association**: **if there are other variables which come in the middle of the process between risk factor and disease.**
3. **Direct (causal) Association**
    1. **One risk factor to one disease causal association**
    2. **Multi-factorial causal association** **(many chronic diseases come under this category)**

Further explanation

Spurious Association: which is mistaken association, the study leads us for association which is not true (for example: Shoe size and reading performance for elementary school children)-

The bigger your shoe size, the better is your reading performance

Is that make sense?

Indeed, there is confounding variable which is age.

In reality,if you are older, your shoe size will be bigger, and your reading performance well be better, so age is what really what associated with performance and not shoe size, but age is related to shoe size. This is what we call confounder, and we should remove its effect before judging about the presence of association, so this is a spurious association.

Association and Causation

Look here to the intersection of the two circles, one circle represents people whom go to the moon, and the other for whom have eaten chicken, the smart guy concluded down here, that chicken makes you fly to the moon, but is that true association? Apparently not! So this is spurious association.

Association or not?

A researcher in his observational study found that the average serum homocysteine among patients of IHD(ischemic heart disease)was15 mcg/dl that higher than normal (while theNormal=10-12 mcg/dl).Can we say that:

- Hyperhomocystenemia causes IHD? Of course not.

But we can Hypothesize that:

- Hyperhomocystenemia **may** have a role in etiology of IHD.

For final proof if there is association or not, 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 (if P-value <0.05, it indicates presence of association, and if is it weak and to which level), 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.

We should go stronger study that provides us with a measure of association, we can go to a case control or coherent or experimental study that give us: relative risk or P-value or incidence, which may give us stronger proof for an association that may be present.

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’.

We usually starts with descriptive studies from which we can generate a new hypothesis, then we need a stronger studies to test this ‘hypothesis testing’: to able to judge if this hypothesis is true or false, accept the hypothesis or reject it.

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”: there is Correct methods used? Good Validity, reliability?is there any Bias,has that been eliminated or reduced?
- **Step 2 has two branches:**

Step 2a: do statistical results indicate association? -by having p value less than 0.05 or 95% CI (confidence interval).

Step 2b: if not significant p value, may be because of low power of the study (smaller sample size)-

Maybe there is an association in the reality, but the small sample size made low power of the study,unable to detect this true association that is found there..so make sure that the power of the study is good enough.

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, that collects the data from more than one study meet on the same thing), rather than straightaway dismissing the 'exposure - outcome' association as non- causal.

Which means we don't jump to conclusions quickly, but first make sure you have enough power of the study, which detects an existing association. Then the next step:

- **Step 3**: if statistically significant –evaluate as to whether this relationship is due to 'indirect relationship' with a third variable (confounder), remember the age as confounder from the previous example? we will explain it soon.
- **Step 4**: if confounder excluded- now test this postulated "causal" relationship on the following **criteria of "causal association"**-association of Hill's, we learned that Hill's criteria that started modern era of epidemiology, we will learn more about them now.

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(or a disease) under investigation, the findings of an association 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.(and its effect can't be removed later on, it becomes intermingled with the study itself)

✓The effect of bias will be an estimate either above or below the true value (>RR or <RR) which means it can lead us to stronger or a weaker association, 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 of bias: 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 differencebetween 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

(remember when we talked about randomized controlled trials? randomization in selecting who goes in to the treatment group or controlled group will avoid us selection bias).

Information bias results from systematic differences in the way data (information) on exposure or outcome are obtained from thevarious study groups (exposed vs non-exposed) (diseased vsnon-diseased). that means the way to collect data about diseased is different from that for non-diseased or healthy, and the same goes for (exposed vs non-exposed).

- **This yields systemic errors in the measurement of exposure or outcome,Which means whether exaggerating the small association or making the association smaller while in the reality is bigger, This will affect the nature of true association (recall bias).**

(Again, the strength of experimental studies, especially randomized controlled trials lies in Blinding, different types of blinding include the participant, physician who observes and collects data from participants, and the analyst who does analysis and doesn't know which participants are exposed to treatment or not exposed but in controlled group, all of that make the study stronger and minimize info bias and selection bias)

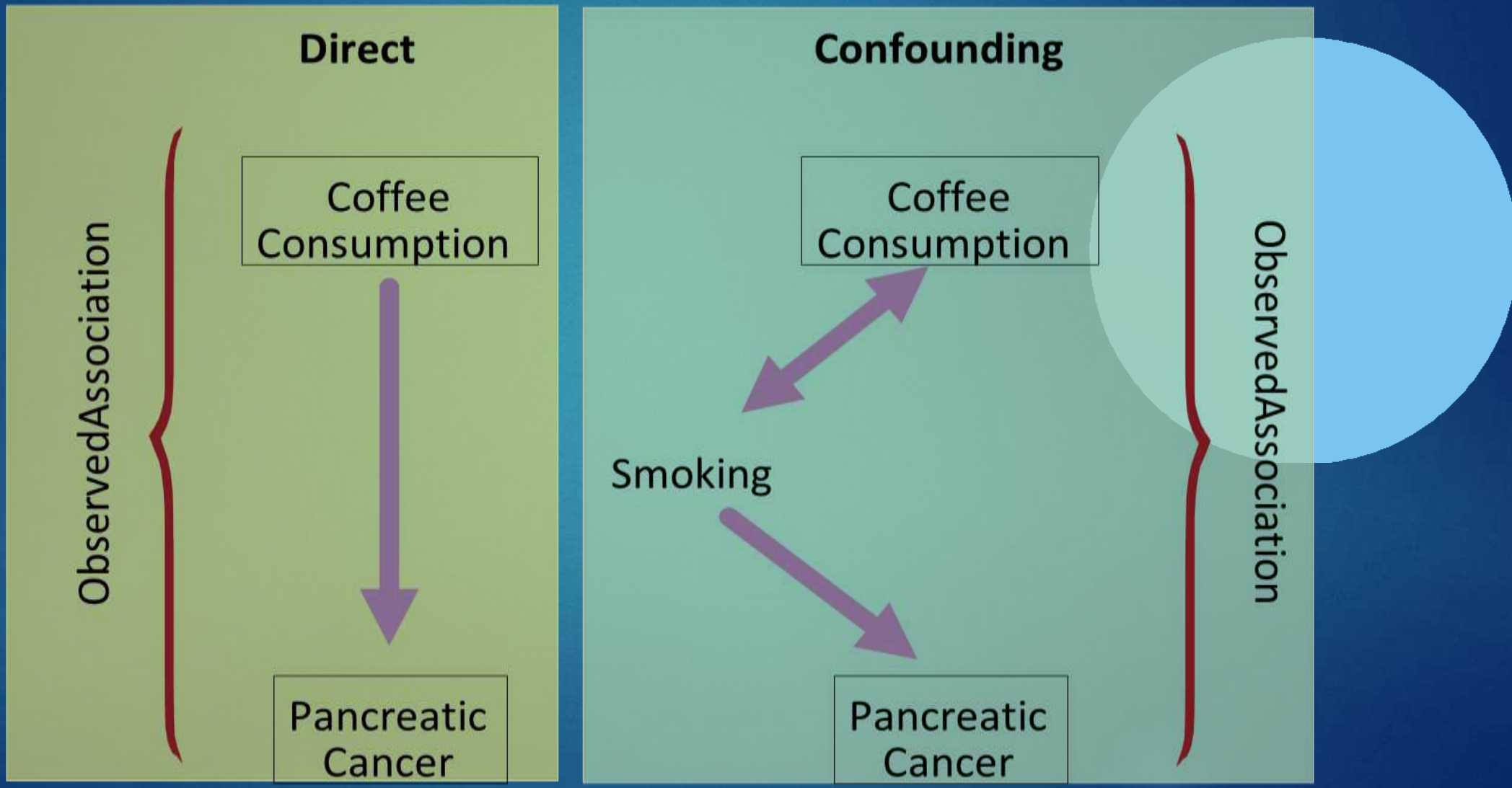
Confounding

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



Further explanation

Confounding could be explained by previous illustration very well, suppose we are examining the relationship between coffee consumption and pancreatic cancer, and we actually observed an association, and we did statistical analysis, and we had P-value less than 0.05, and we have high relative risk(RR), and everything make sense that there is association..

But we still have to stop and think: is that association real?

If we get deeper, we find that coffee consumption is associated with smoking, and smoking is a risk factor for pancreatic cancer, and smoking isn't in causal pathway between coffee consumption and pancreatic cancer in physiological or pathological pathway..but we see that people who drink coffee, smoke more, and since smoking is established risk factor, so it may be the real cause! Although it seems the higher coffee consumption, the higher chances of pancreatic cancer.

How can we determine? During analysis stage to rule out the effect of smoking and compensate for it, and see if we remove smoking, would still there be an association between coffee consumption and pancreatic cancer?

So we have to dig deep into our mind, and see which variables could be confounders for a relationship of coffee consumption and pancreatic cancer, and smoking cold be right in this case.

Confounding

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

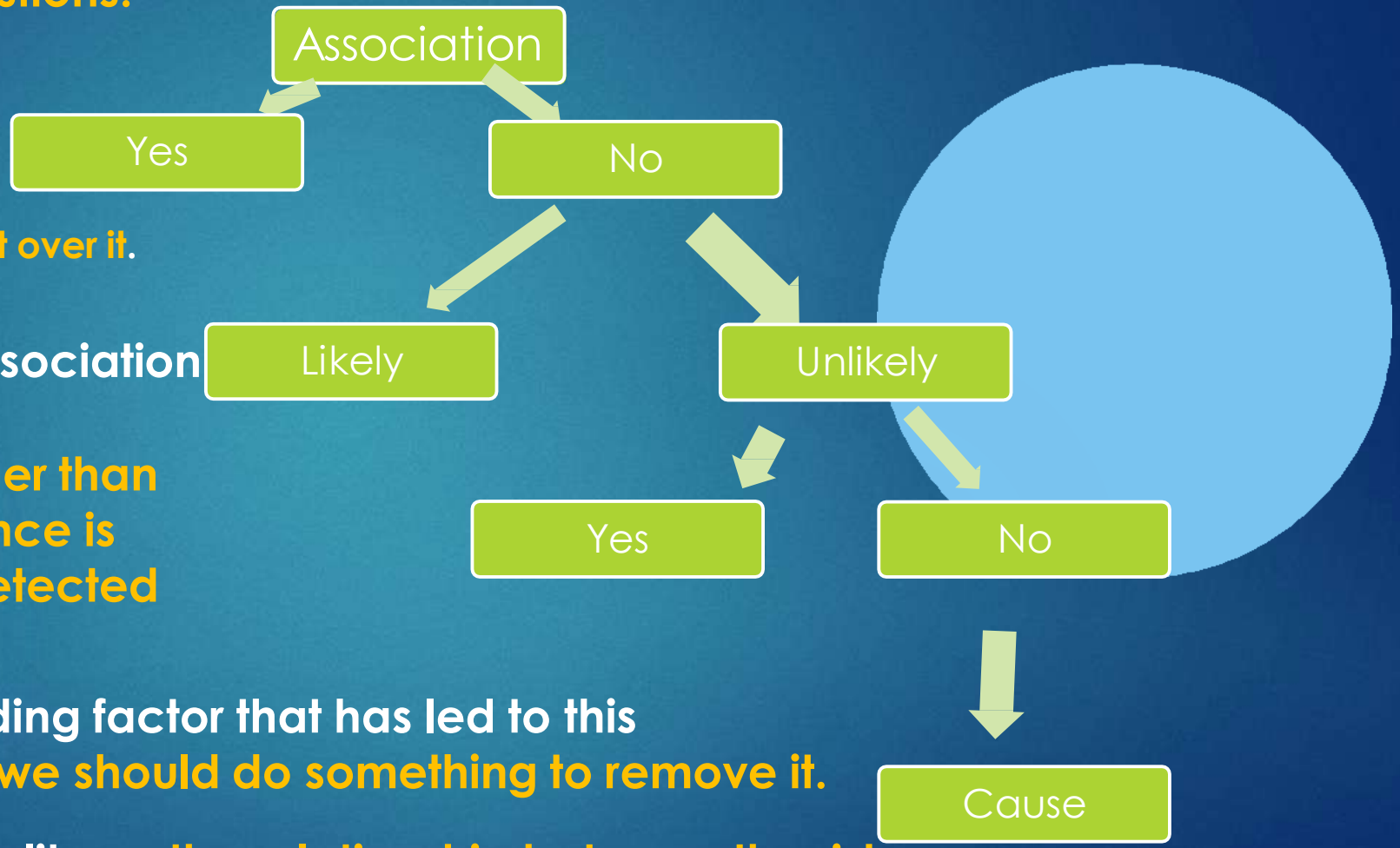
Confounding is different from bias, may lead to errors in the conclusion of a study, but, when confounding variables are known, the effect may be fixed or removed (corrected, accounted for, controlled for) and finding out the pure to true association between the risk factor and the disease of interest.

-Controlling of confounding(removing its effect) at the design stage by: restriction ,matching(match every case with very similar control as in case control study), And randomization(that we do it in randomized control trials).

If we discovered the confounding factor later on, we can still control it in analysis stage:

-Controlling of confounding at the analysis stage: stratification(for example if the suspected confounder is age, we can make 3 strata of age, higher and lower and in between , and see if there is any difference),multivariate analysis(very common and it depends on removing the confounders to know the cause), and standardization(according confounding variables used during the analysis stage).

To go From Association to Causation, we should answer these questions:



1)Bias in selection or measurement?

-If yes we should do something to get over it.

2) Chance is the reason of association that we observe?

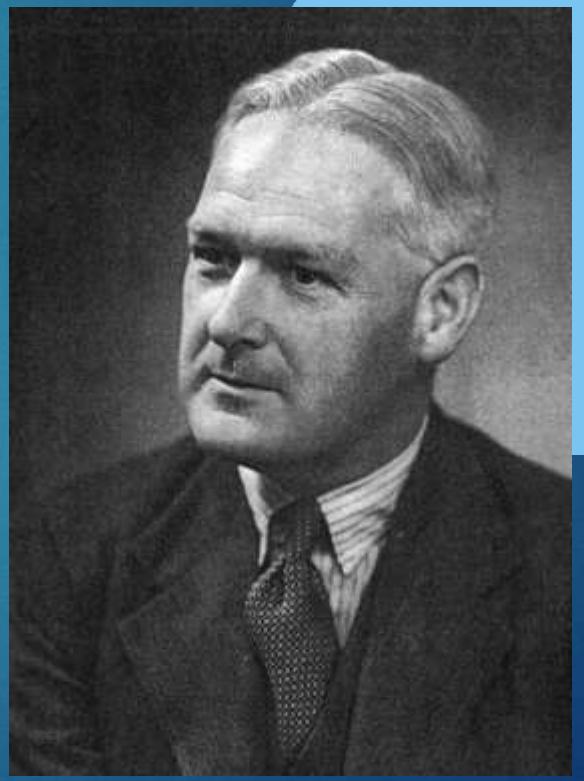
If it is likely, so p-value is higher than 0.05, but if less than 0.05,chance is unlikely to be the cause of detected association.

3) Could there be a Confounding factor that has led to this apparent association? If yes, we should do something to remove it.

4)Finally,Hill's criteria for Causality on the relationship between the risk factor and disease, and see which of these criteria does apply to the situation that we have.

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?*
  - *(to go from association into judging a causal association)*
  - *→ so he came up with his criteria .*



Guidelines for judging whether an association is causal

Sir Austin Bradford Hill criteria

- **Most Important criteria (obligatory to be found)**

1. **Temporality:** cause or risk factor precedes effect, this should be confirmed.
2. **Strength of association**: large relative risk (the larger relative risk, the stronger the association), this factor always should be present.
3. **Consistency**: repeatedly observed by different. persons, in different places, circumstances, and times. They all observed such an association between this risk factor and the disease.

Guidelines for judging whether an association is causal

Additional supportive criteria (extra support)

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. (isn't obligatory, because our knowledge about the biology is limited, and things may be discovered in the future).
6. **Experimental evidence. (experimental studies have provided more evidence about this risk factor and the disease)**
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??????

(this question is included in the final exam, so please go and read the introduction, you will enjoy it ^_^)