Measures of Association in Epidemiology

Dr. Sireen Alkhaldi
Community Medicine
Faculty of Medicine, The University of Jordan

In this character:

There is a student who is going to the elections, he's counting how many of the girls & boys are going to vote for him →21% of the boys & 30% of the girls support him so he thinks he will get 51% of the votes →he thinks he will win

DO you think he has got it right ? 51% of the votes? Of course not , he is mistaken(his fault was that he misinterpreting the results: he has got accurate numbers (21% & 30%) ,but that doesn't mean that 51% of all students will vote for him.

What we should learn is that :

-it is very important to know how to read & understand the numbers

-how to interpret statistics (that you collect) in the right way in order to get the right meaning out of them

the main Measures of Association in Epidemiology

- Chi square
- Odds Ratio
- Relative Risk or Risk Ratio
- Attributable Risk

2X2 Table (contingency table)

		Disease		
		Yes (+)	No (-)	Total
Exposure	Yes (+)	a	b	a+b
	No (-)	c	d	c+d
	Total	a+c	b+d	a+b+c+d

The main tool that we are going to use is contingency table (2X2 table) :

A table that contains **4 cells** (a,b,c,d) on the **left:** the exposure (always on the left)

Up: the disease status

yes(+) →has the disease

No(-) →does not have the disease

On the left :

yes(+) → exposed

No(-) →non-exposed

We also need to read the subtotals :

The exposed are a+b

non-exposed are c+d

Those who have the disease a+c

Don't have the disease b+d

a+b+c+d: the grand total for all the participants

We are going to use this table to calculate all the measures of association

Cells:

Cell A= Exposed, and diseased

Cell B= Exposed, Not diseased

Cell C= Not exposed, diseased

Cell D= Not exposed, Not diseased

A+B+C+D=Total

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

Practice drawing the table : put the exposure status on the left, disease status up, make sure you understand the cells & the subtotals

Totals

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

Marginal totals

a+b= Exposed

c+d= Non-exposed

a+c= Diseased

b+d= Non-diseased

Grand total

n = a+b+c+d

1. Chi-square in Cross-sectional studies to look for an association between a certain risk factor & a disease

Chi-square tests whether there is an association between two categorical variables.

If we have a continuous variable (age for example) we have to make it into a categorical variable like dichotomous variable age could become over 45 and under 45 so we can use it in 2*2 table and do chi-square

For a 2X2, table: $X^2 = \frac{n(ad - bc) - n/2)^2}{(a+b)(a+c)(c+d)(b+d)}$ **equation for calculating chi square from 2*2**

Now a day NO ONE calculate chi-square using this equation manually, we usually use a software like SPSS or excel to calculate chi-square just by providing the cells / there are too many ways to calculate to automatically not manually

Extra information (not mentioned in this lecture) : Dichotomous : a division into two (only two) example sex (male & female) we will talk about it more in biostatistics

the importanat thing is the P-value, first we will get a number for the chi-square(we don't get a lot of meaning from that., but we look for a corresponding p-value if<0.05 (<0.05 that is usually set p-value for statistical significance) so if it is less than 0.05 we say that there is a significant association between the risk factor (exposure)and the disease

So If the calculated chi-square value is greater than the critical value or P<0.05, we say that there is a significant association between the risk factor and the disease

Chi-square statistic tells only whether there is association or no association. It doesn't tell us how strong an association is we cannot know that but this information will be available when we calculate odds ratio or risk ratio, it can give magnitude of association (how strong it is or whether it is a weak association

2. Relative risk (RR) or Risk Ratio (RR) (the same symbol) In a cohort study using contingency table

RR: The estimation of disease risk associated with exposure (indication of strength of association)→how strong or weak the association is if it was present

RR Expresses risk of developing a disease in exposed group (a + b) as compared to non-exposed group (c + d)

We use this equation after filling the cells correctly (exposed or not/ has the disease or not)

$$RR = \frac{\text{Incidence (risk) among exposed}}{\text{Incidence (risk) among non-exposed}}$$

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

Analysis in Cohort studies

In a Cohort Study, we can calculate Incidence.

So, Relative Risk can be obtained from a cohort study.

Risk = incidence & relative risk = incidence among exposed compared to the non-exposed

Example :

Cigarette smoking (Exposure)	*Disease (with Ca lung)*	*No Disease (without Ca lung)*	*Total*
Yes	*70 (a)*	*6930 (b)*	*7000 (a+b)*
No	*3(c)*	*2997(d)*	*3000 (c+d)*

We are going to calculate relative risk, the exposure of interest here is cigarette smoking we have 2 groups: smoker & non-smokers , the disease is lung cancer, we have 2 groups : with lung cancer & without lung cancer

RR in a Cohort Study

Incidence rates :

Risk among exposed (smokers) = 70/7000 = 10 / 1000.

Risk among non-exposed(non smokers) = 3/3000 = 1 / 1000.

$$RR = \frac{\text{Risk (Incidence) among exposed}}{\text{Risk (Incidence) among non exposed.}}$$

RR = 10/1=10

Interpret (report/give the meaning) RR :The exposed have 10 times the risk of developing the disease when compared to non-exposed smokers are 10 times at risk of developing lung cancer compared to non smokers

Interpretation of relative risk

What does a RR of 2 mean?

Risk in exposed=2X Risk in non-exposed

Thus a relative risk of 2 means the exposed group is two times at a higher risk of developing the disease when compared to non-exposed

Strength of association

In general strength of association can be considered (interpreted) as:

High association if RR≥3

Moderate if RR is between 1.5 & 2.9

Weak association if RR is between 1.2 & 1.4

No association exists if RR is 1(since it is a ratio)

Negative association (protective effect) if RR <1

Example :

RR=0.7 → that means that the incidence of disease among exposed is less than the disease among non-exposed, that would be a protective effect of a risk factor so it is a productive factor (negative association)→ present of the exposure (risk factor) prevents the disease occurrence among exposed

3. ODDS RATIO (OR)

Odds Ratio (OR) is a measure of the strength of the association between risk factor & outcome (disease).

✓The odds ratio is the cross product of the entries (a,b,c,d) in table.

Cross product of the table = (a*d)/(b*c)

✓ OR can be calculated in case-control studies instead of RR since case control studies does not provide incidents like a cohort study does , we cannot calculate relative risk from a case control study

❑ OR can be a good estimate of RR something that gives an estimate related to RR which is very close to it and can be calculated from case control study

Odds ratio (OR)

Odds Ratio can be a good estimate of RR.

Odds ratio is the ratio of odds of exposure among diseased to odds of exposure among non-diseased

$$OR = \frac{\text{Odds of exposure among diseased}}{\text{Odds of exposure among non-diseased}}$$

$$= (a/c)/(b/d) = ad/bc$$ **(cross product of the table calculated in a simpler way than RR)**

Interpretation of OR is the same as that of RR

Interpretation of OR is the same as that of RR

In general strength of association can be considered (interpreted) as:

High(strong) association if OR≥3

No association exists if OR is 1

Negative association (protective effect) if OR <1

Odds ratio…

RR can be best estimated by OR if the following conditions are fulfilled:

1. Controls are representative of general population
2. Selected cases are representative of all cases in the population
3. The disease is rare (the more rare the disease is the more close would the OR be to the RR)

ANALYSIS in case-control studies

Estimation of disease risk associated with exposure (odds ratio)

OR in a case-control study

an example about an exposure (smoking) , in the left (smokers & non-smokers) / the disease: lung cancer (we have cases with lung cancer & controls without lung cancer)

we have cells (a,b,c and d)→it is very important if you are giving information about number of smokers , non-smokers, how many cases . That you can create your own 2*2 table in the correct way and fill the cells correctly (this is the basis starting from there it is very easy to calculate odds ratio

	Cases (with Ca lung)	*Controls (without Ca lung)*
Smokers	*33 (a)*	*55 (b)*
Non smokers	*2 (c)*	*27 (d)*
Total	*35 (a+c)*	*82 (b+d)*

OR =cross product of the table =(33*27)/(55*2)=8.1

OR in Case-control studies

Odds ratio is a key parameter in the analysis of Case-control studies

	Disease: Cases (Lung Ca)	Disease: Controls (No Lung Ca)
Smokers	***a (33)***	***b (55)***
Not smokers	***c (2)***	***d (27)***

$$\text{Odds Ratio} = ad/bc = \frac{33 \times 27}{55 \times 2} = 8.1$$

Interpret it :the odds of disease among the exposed is 8 higher compared to non-exposed , so the exposure increase the odds (the probability)of getting lung cancer compared to non-exposed

Attributable Risk (AR)

AR indicates how much of the risk is due to (attributable to) the exposure.

AR Quantifies the excess risk in the exposed that can be attributable to the exposure, by removing the risk of the disease that occurred due to other causes.

AR= Risk (incidence) in exposed (-) Risk (incidence) in non-exposed

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

AR= [a/(a+b)] - [c/(c+d)]

Attributable risk is also called **risk difference** & that all refers

To the excess risk provided by this exposure understudy

Attributable risk percent (AR%)

Estimates the proportion of disease among the exposed that is attributable to the exposure.

Very similar to calculating AR with little difference

$$AR\% = \frac{(\text{Risk in exposed} - \text{Risk in non-exposed})}{\text{Risk in exposed}} \times 100\%$$

(to get the percent)

(to get proportion)

Go back to the example in slides 11&12

$$AR\% = \frac{10 - 1}{10} \times 100\% = 90\%$$

(as in the previous Cohort study example)

(10 was the incidence among exposed 1 among non)

90% of the lung cancer among smokers was due to their smoking.(interpret it)

- This suggests the amount of disease that might be eliminated if smoking could be controlled or eliminated. Is 90% of lung cancer

Possible outcomes in studying the relationship between exposure & disease

1. No association

   RR=1

   AR=0 (the difference between the incidence among exposed & non exposed)

2. Positive association

   RR>1

   AR>0

3. Negative association

   RR<1 (fraction)

   AR<0 (Negative)

Risk Vs Preventive factors

A ***risk factor*** is any factor positively associated with a disease (RR>1). It is associated with an increased occurrence of a disease

A ***preventive factor*** is any factor negatively associated with a disease (RR<1). It is associated with a decreased occurrence of a disease.

Risk and preventive factors ***may (not)*** be amenable to change (e.g. Smoking, age) PH is interested in factors that we can change (we can do something about/delete) for example we cannot change the person genetic makeup or change a person's age, but we can change eating habits or smoking)

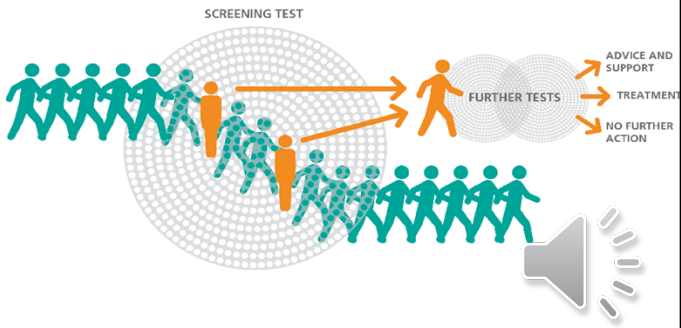
Population Screening

This topic that not related to measures of association , but pop. Screening that something we have to know about epidemiology.

Screening is “application of a test to **asymptomatic** people to detect **occult disease[subclinical disease]** or a precursor state” (Alan Morrison, *Screening in Chronic Disease*, 1985)

- Immediate objective of a screening test – to classify people as being likely or unlikely of having the disease.
- Ultimate objective: to reduce mortality and morbidity
- Many countries have national screening programs for various diseases Specially for diseases that are becoming common

if we want to wait until people get diseased it will kill people or become very hard treat or cost a lot to treat or make great disability if we wait until the symptoms show people→ so we must try to detect it by screening of population In different stages.



Population Screening

Why do we have cancer screening in very different ways? Because cancer is a very chronic disease that takes very long time years to progress until it becomes discoverable in a clinical setting. So we have a chance to detect and discover the cancer before getting bigger and able to be detectable in clinical setting.

Familiar examples of the use of population screening are **cancer** screening (e.g., Pap smears to detect cervical neoplasia, mammography and physical breast exam to detect breast cancer, PSA to detect early prostate cancer, and fecal occult blood testing to detect colorectal cancer or adenomas [which are the most common types of cancer and have very high probability of detecting many cases in preclinical stages years before it becomes big enough to be detected by physician]).

Screening programs for **hypertension and diabetes** to prevent complications

Screening of **newborns** for phenylketonuria (PKU) to prevent mental retardation.

PCR test for Covid-19, using thermometer to detect fever).

More details about previous slide→

1-Screening programs for **hypertension and diabetes** to prevent complications→ a country like Jordan, we have a large proportion of adult people Who will develop diabetes and Hypertension. so if you can do screening for all they go to physician will always do the test of diabetes and just get BP measured to see if there is hypertension → we can detect many of the cases in early stages and start prevention before complications happened.

2-Screening of **newborns** for phenylketonuria (PKU) to prevent mental retardation→ very simple needle in the heal of the baby after being born , a small drop of blood can be taken to do the test and see if the baby has PKU deficiency or not, because if we don't detect it will cause mental retardation, but early detectable will prevent this occurring.

Screening test

A screening test must be accurate. There are two major dimensions to accuracy:

1. Reliability (consistent): get same result if we repeat the test again (However, a test can be reliable but still give an incorrect result).

2. Validity: give the correct result (reflect the true state).

There are two desirable properties for a screening test related to validity:

Sensitive: The ability of test to correctly classify cases as positive.

Specificity: The ability of test classify non-cases as negative.

Note that sensitivity and specificity are both probabilities of correct classification of cases and non-cases by screening test.

Reliability

Reliability does not ensure validity, but lack of reliability constrains and jeopardizes validity.

A test that is unreliable cannot be valid. The opposite is not true, however (a reliable test may or may not be valid).

Even if test-retest agreement is very high (for example, 100%), the test could simply be consistently incorrect.

e.g. measurement of blood pressure using sphygmomanometer, temperature using thermometer.

Validity

1. Sensitivity of the test: the probability of correctly classifying someone who has the disease (case).

Probability (proportion) of correct classification of cases (by screening test) :

Cases found by the test as(positive) + / all cases

2. Specificity of the test: the probability of correctly classifying someone without the disease (healthy, non-case).

Probability (proportion) of correct classification of **non-cases:**

Non-cases found by the test as

the number of (non-cases/negative) -/ all non-cases