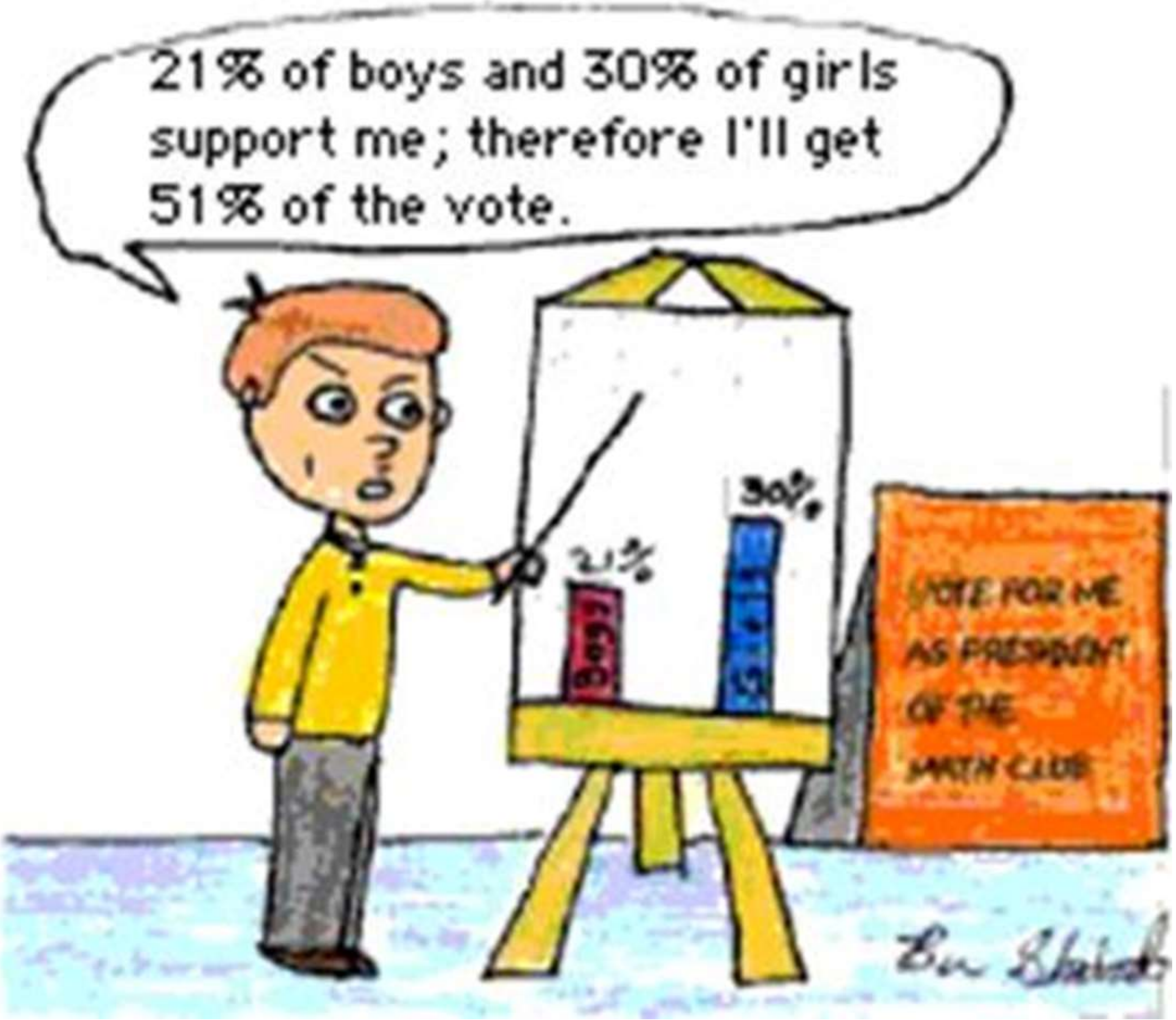
# Measures of Association in Epidemiology

**Dr. Sireen Alkhaldi**

**Community Medicine 2022/ 2023**

**Faculty of Medicine, The University of Jordan**


21% of boys and 30% of girls support me; therefore I'll get 51% of the vote.
21%
30%

# Measures of Association in Epidemiology

- Chi square
- Odds Ratio (OR)
- Relative Risk or Risk Ratio (RR)
- Attributable Risk (AR)

# 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 |

# Cells

| 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 |

A= Exposed, and diseased

B= Exposed, Not diseased

C= Not exposed, diseased

D= Not exposed, Not diseased

A+B+C+D=Total

# 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

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

For a 2X2, table:

$$X^2 = \frac{n\,(ad - bc) - n/2)^2}{(a+b)(a+c)(c+d)(b+d)}$$

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 (usually calculated using software like SPSS or excel).

Chi-square statistic tells only whether there is association. It doesn't tell us how strong an association is.

# 2. Relative risk (RR) or Risk Ratio (RR) In a cohort study

RR: The estimation of disease risk associated with exposure (strength of association)

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

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

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

| 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.

| 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) |

# 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**

The exposed have 10 times the risk of developing the disease when compared to non-exposed

# 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 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

Negative association (protective effect) if RR <1

# 3. ODDS RATIO (OR)

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

✓The odds ratio is the cross product of the entries in table.

✓ OR can be calculated in case-control studies instead of RR

2 by 2 table

| | Diseased - Cases | Non-diseased – Controls | Total |
|---|---|---|---|
| Exposed | A | B | A+B |
| Non-exposed | C | D | C+D |
| Total | A+C | B+D | A+B+C+D |

# 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

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

$$= (a/c)/(b/d) = ad/bc$$

**Interpretation of OR is the same as that of RR**

# 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
3. The disease is rare

# ANALYSIS in case-control studies

## Estimation of disease risk associated with exposure ( odds ratio )

# OR in a case-control study

| | 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 in Case-control studies

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

Disease

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

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

# Attributable Risk (AR)

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

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**

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

Attributable risk is also called risk difference.

# Attributable risk percent (AR%)

Estimates the proportion of disease among the exposed that is attributable to the exposure (example of smoking and lung C).

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

$$AR\% = \frac{10 - 1}{10} \times 100\% = 90\%$$ (as in the previous Cohort study example)

- **90% of the lung cancer among smokers was due to their smoking.**
- **This suggests the amount of disease that might be eliminated if smoking could be controlled or eliminated.**

# Possible outcomes in studying the relationship between exposure & disease

1. No association
   - RR=1
   - AR=0
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

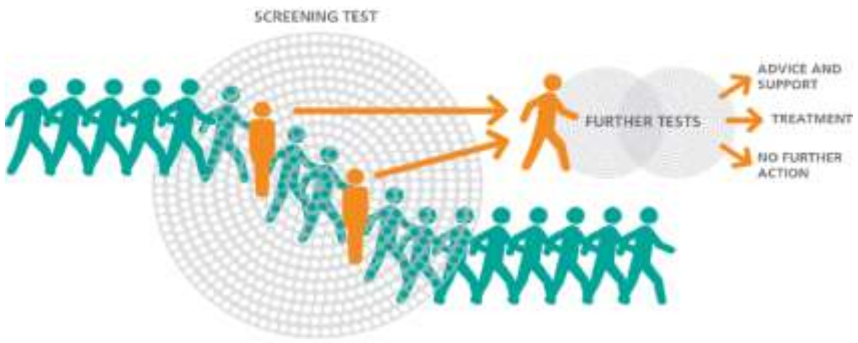
A ***preventive factor*** is any factor negatively associated with a disease (RR<1). It is associated with a decreased occurrence of a disease.

Risk and preventive factors ***may (not)*** be amenable to change (e.g. Smoking, age)

# Population Screening

Screening is “application of a test to **asymptomatic** people to detect **occult 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.



# Population Screening

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).

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.

# Screening test

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

1. Reliability (consistent): get same result if we repeat it (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 to correctly classify cases as positive.

Specificity: The ability classify non-cases as negative.

Note that sensitivity and specificity are both probabilities of correct classification.

# 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 **:**

**Cases found by the test as + / 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 - / all non-cases**

Status of person
according to "gold standard"

| Result from screening test | | Has the condition | Does not have the condition | |
|---|---|---|---|---|
| | Positive | a True positive | b False positive | ← Row entries for determining **positive predictive value** |
| | Negative | c False negative | d True negative | ← Row entries for determining **negative predictive value** |
| | | ↑ Column entries for determining **sensitivity** | ↑ Column entries for determining **specificity** | |

# Sensitivity and Specificity

$$Sensitivity = \frac{Number\ of\ true\ positives}{(Number\ of\ true\ positives\ +\ Number\ of\ false\ negatives)}$$

$$= \frac{Number\ of\ true\ positives}{Total\ number\ of\ individuals\ with\ the\ illness}$$

$$Specificity = \frac{Number\ of\ true\ negatives}{(Number\ of\ true\ negatives\ +\ number\ of\ false\ positives)}$$

$$= \frac{Number\ of\ true\ negatives}{Total\ number\ of\ individuals\ without\ the\ illness}$$

## True Disease Status

| Screening Test Results | | Cases | Non-cases | |
|---|---|---|---|---|
| | Positive | True positive (a) | False positive (b) | a + b |
| | Negative | False negative (c) | True negative (d) | c + d |
| | | **a + c** | **b + d** | |

$$\textbf{Sensitivity} = \frac{\text{True positives}}{\text{All cases}} = \frac{a}{a + c}$$

$$\textbf{Specificity} = \frac{\text{True negatives}}{\text{All non-cases}} = \frac{d}{b + d}$$

## True Disease Status

| Screening Test Results | | Cases | Non-cases | |
|---|---|---|---|---|
| | Positive | 140 (a) | 1,000 (b) | 1,140 |
| | Negative | 60 (c) | 19,000 (d) | 19,060 |
| | | 200 | 20,000 | |

$$\textbf{Sensitivity} = \frac{\text{True positives}}{\text{All cases}} = \frac{140}{200} = 70\%$$

$$\textbf{Specificity} = \frac{\text{True negatives}}{\text{All non-cases}} = \frac{19{,}000}{20{,}000} = 95\%$$

# Interpreting test results: predictive value

Probability (proportion) **of those tested** who are **correctly classified by the test:**

**Positive predictive value** = Cases identified / all positive tests

**Negative predictive value** = Non-cases identified / all negative tests

## True Disease Status

| | | Cases | Non-cases | |
|---|---|---|---|---|
| Screening Test Results | Positive | True positive (a) | False positive (b) | **a + b** |
| | Negative | False negative (c) | True negative (d) | **c + d** |
| | | a + c | b + d | |

$$\textbf{PPV} = \frac{\text{True positives}}{\text{All positives}} = \frac{a}{a + b}$$

$$\textbf{NPV} = \frac{\text{True negatives}}{\text{All negatives}} = \frac{d}{c + d}$$

True Disease Status

| Screening Test Results | | Cases | Non-cases | |
|---|---|---|---|---|
| | Positive | 140 (a) | 1,000 (b) | 1,140 |
| | Negative | 60 (c) | 19,000 (d) | 19,060 |
| | | 200 | 20,000 | |

$$\textbf{PPV} = \frac{\text{True positives}}{\text{All positives}} = \frac{140}{1{,}140} = 12.3\%$$

$$\textbf{NPV} = \frac{\text{True negatives}}{\text{All negatives}} = \frac{19{,}000}{19{,}060} = 99.7\%$$