Probability in Epidemiology

DR.ALAA MOHAMMED

3RD LEVEL, 2^{ND} SEMESTER, BIOMEDICAL INSTRUMENTATION AND BIOMECHANICS BRANCHES, BIOMEDICAL ENG. DEPARTMENT

Epidemiology

Epidemiology and **biostatistics** are the basic sciences of public health

Public health investigations use quantitative methods, which combine the two disciplines of epidemiology and biostatistics

Epidemiology is defined as the study of the distribution and determinants of health, disease, or injury in human populations and the application of this study to the control of health problems

Epidemiology represents the understanding of disease development and the methods used to uncover the aetiology, progression, and treatment of the disease

Epidemiology (cont.)

Epidemiology is the branch of medicine which deals with the incidence, distribution, and possible control of diseases and other factors relating to health.

Epidemiology is a strategy for the study of factors relating to the etiology, prevention, and control of disease, maintenance and medical care in human populations, and away of studying a health problem.

Epidemiology can help us identify and understand the factors that influence the emergence, severity, and consequences of health problems

Uses of Epidemiology

1. To make a community diagnosis
2. To describe magnitudes of disease
3. To know causation of disease
4. To know natural history of a disease
5. To monitor continuously over a period of time the change of health in a community. (for example, the effect of a vaccination program, health education and nutritional supplementation).
6. To practice surveillance for a specific disease in order to be able to act quickly and so cut short any outbreak (for example cholera).
7. Description of health status in population
8. Health planning and identifying priorities
9. Evaluation of intervention
10. Epidemiology helps to identify and describe health problems in a community (for example, the prevalence of anemia, or the nutrition status of children).

Probability in Epidemiology

❖ Definition of probability

The probability of an event is the frequency with which it is expected to occur in a long series of trials.

Simple probability. $P(A)$**.** The probability that an event (say, A) will occur.

Joint probability. $P(A\ and\ B)$. $P(A\ \cap\ B)$. The probability of events A and B occurring together.

Conditional probability. $P(A|B)$, read "the probability of A given B." The probability that event A will occur given event B has occurred.

❖ Rules of Probability

1. The probability of an event (say, event A) cannot be less than zero or greater than one.

E.g., The probability that you will pass this course cannot be 120%.

So, $\mathbf{0} \leq \boldsymbol{P(A)} \leq \mathbf{1}$

2. The sum of the probabilities of all possible outcomes (events) of a process (or, experiment) must equal one.

So, $\sum \boldsymbol{Pi} = \mathbf{1}\ \boldsymbol{or}\ \boldsymbol{P(A)} + \boldsymbol{P(A')} = \mathbf{1}$ **[**$\boldsymbol{A'}$ **means *transpose of A*.]**

Rules of Probability (cont.)

3. Rules of addition.

a. $\boldsymbol{P(A\ or\ B)\ =\ P(A\ U\ B)\ =\ P(A)\ +\ P(B)}$

if events A and B are mutually exclusive.

Two events are *mutually exclusive* if they cannot occur together. E.g., male or female; heads or tails.

b. In general, **P(A or B) = P(A) + P(B) - P(A and B)**

- This is the *general formula* for addition of probabilities, for any two events A and B.
- **P(A and B)** is the **joint probability** of A and B occurring together and is equal to zero if they are mutually exclusive (i.e., if they cannot occur together).

Rules of Probability (cont.)

4. **Rules of multiplication** are used for determining joint probabilities, the probability that events A and B will occur together.

$$\boldsymbol{P(A\ and\ B) = P(A \cap B) = P(A) \cdot P(B)}$$

if events A and B are independent. Events A and B are *independent* if knowledge of the occurrence of B has no effect on the probability that A will occur.

❖ Probability distribution

A probability distribution describes the true relative frequency of all possible values of a random variables

General addition principle (for any event)

$$\Pr(\boldsymbol{A\ or\ B}) = \Pr(\boldsymbol{A}) + \Pr(\boldsymbol{B}) - \Pr(\boldsymbol{A\ and\ B})$$

❖ Probability in Diagnostic Testing

a) **Prevalence ("existing")** The point prevalence of disease is the proportion of individuals in a population that has disease at a single point in time (point), regardless of the duration of time that the individual might have had the disease.

Prevalence is NOT a probability

b) **Incidence ("new")** The incidence of disease is the probability an individual who did not previously have disease will develop the disease over a specified time period.

Probability in Diagnostic Testing (cont.)

c) Sensitivity, Specificity

Sensitivity and *specificity* are terms used to evaluate a clinical test.

Positive and negative predictive values are useful when considering the value of a test to a clinician.

They are dependent on the prevalence of the disease in the population of interest.

The ideas of *sensitivity, specificity, predictive value of a positive test, and predictive value of a negative test* are most easily understood by using data in the form of a 2x2 table (1):

Probability in Diagnostic Testing (cont.)

In this table, each of a total of $(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}+\boldsymbol{d})$ individuals are cross-classified according to their values on two variables:

disease (present or absent) and test result (positive or negative)

It is assumed that a positive test result is suggestive of the presence of disease.

Table (1)

		Disease Status		
		Present	**Absent**	
Test Result	**Positive**	a	b	a + b
	Negative	c	d	c + d
		a + c	b + d	a + b + c + d

Sensitivity and Specificity

The counts have the following meanings:

- **a** = number of individuals who test positive AND have disease
- **b** = number of individuals who test positive AND do NOT have disease
- **c** = number of individuals who test negative AND have disease
- **d** = number of individuals who test negative AND do NOT have disease
- **(a+b+c+d)** = total number of individuals, regardless of test results or disease status

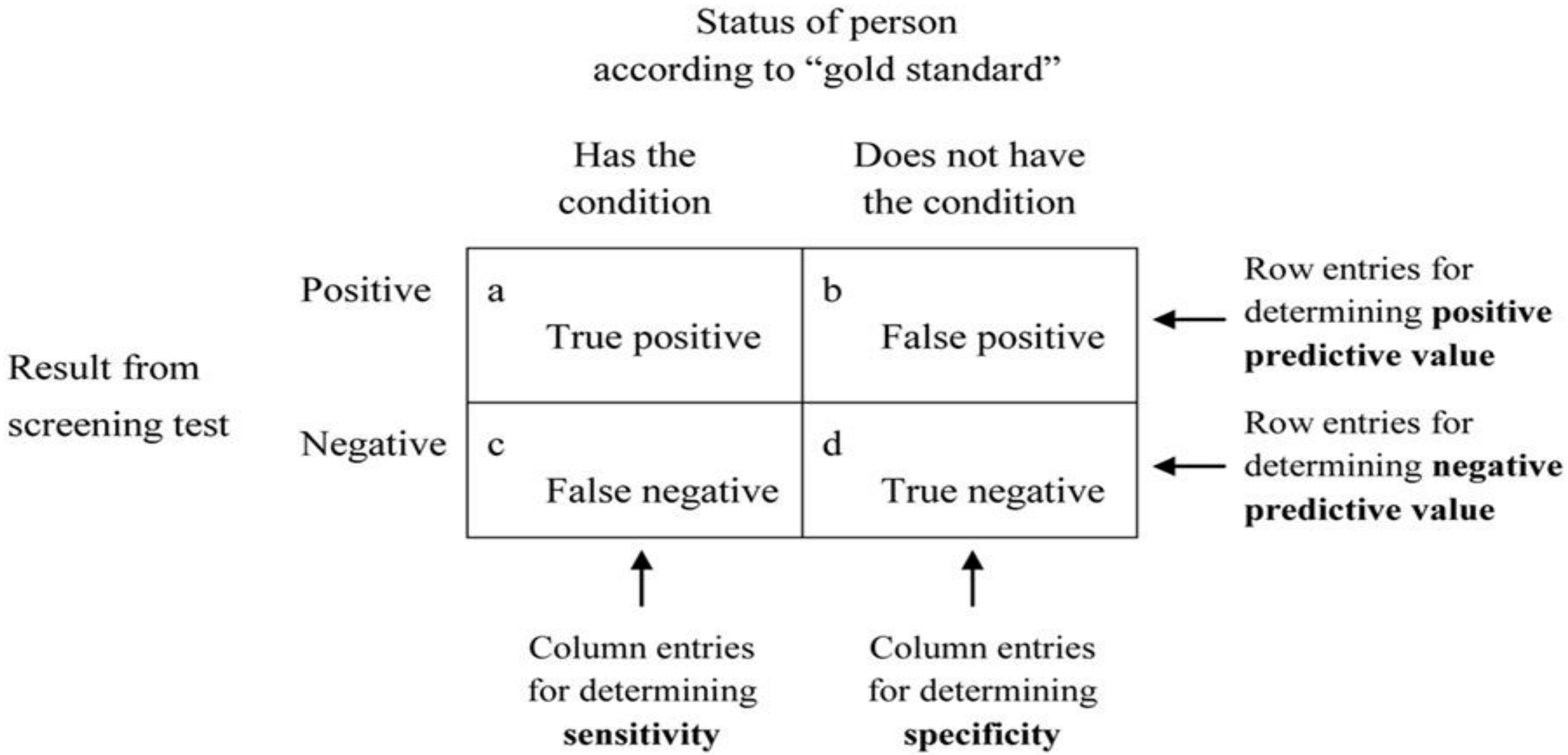
		Disease Status		
		Present	Absent	
Test Result	Positive	a	b	a + b
	Negative	c	d	c + d
		a + c	b + d	a + b + c + d

Sensitivity and Specificity (cont.)

- **(b + d) =** total number of individuals who do NOT have disease, regardless of their test outcomes
- **(a + c) =** total number of individuals who DO have disease, regardless of their test outcomes
- **(a + b) =** total number of individuals who have a POSITIVE test result, regardless of their disease status.
- **(c + d) =** total number of individuals who have a NEGATIVE test result, regardless of their disease status.

		Disease Status		
		Present	**Absent**	
Test Result	**Positive**	a	b	a + b
	Negative	c	d	c + d
		a + c	b + d	a + b + c + d

Sensitivity and Specificity (cont.)



Sensitivity

The sensitivity of a diagnostic test is expressed as the probability (as a percentage) that a sample tests positive given that the patient has the disease.

The following equation is used to calculate a test's sensitivity:

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

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

Specificity

The specificity of a test is expressed as the probability (as a percentage) that a test returns a negative result given that the that patient does not have the disease.

The following equation is used to calculate a test's specificity:

$$\boldsymbol{Specifity = \frac{Number\ of\ true\ negatives}{(Number\ of\ true\ negatives + Number\ of\ false\ pasitive)}}$$

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

Predictive Values (Positive Test and Negative Test)

d) Predictive Value Positive, Negative

“For the person who is known to test positive, what are the chances that he or she truly has disease?".

This is the idea of “predictive value positive test”

$$\boldsymbol{Predictive\ value\ positive\ =\frac{a}{a+b}}$$

"For the person who is known to test negative, what are the chances that he or she is truly disease free?".

This is the idea of “predictive value negative test”

$$\boldsymbol{Predictive\ value\ negative=\frac{d}{c+d}}$$

Predictive Value Positive Test

The positive predictive value is calculated using the following equation:

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

$$PPV = \frac{Number\ of\ true\ positives}{Number\ of\ Samples\ that\ tested\ positive}$$

Predictive Value Negative Test

The negative predictive value is calculated using the following equation:

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

$$NPV = \frac{Number\ of\ true\ negatives}{Number\ of\ samples\ that\ tested\ negative}$$

Prevalence & Incidence

Prevalence is define the number of affected persons divided on the persons in population

Prevalence represents the measure of disease

Incidence is define the number of new cases of a disease in specific time period divided on person at risk at the same time period then multiply the result by 100 or 10000 based on the number of population

Incidence represents the measure of disease risks

Prevalence looks at existing cases, while incidence looks at new cases

Prevalence & Incidence (cont.)

$$\textbf{Prevalence} = \frac{\boldsymbol{no.of\ cases\ (diseased)}}{\boldsymbol{total\ population}} \times \mathbf{100}$$

$$\boldsymbol{Incidence} = \frac{\boldsymbol{new\ cases}}{\boldsymbol{total\ population}}$$

Example 1

You have a new diagnostic test that you want to evaluate. You have a panel of validation samples where you know for certain whether they are definitely from diseased or healthy individuals for the condition you are testing for.

Your sample panel consists of 150 positives and 400 negatives. compare your results to their known disease status and find:

- True positives (test result positive and is genuinely (true) positive) = 144
- False positive (test result positive but is actually negative) = 12
- True negatives (test result negative and is genuinely negative) = 388
- False negative (test result negative but is actually positive) = 6

Example 1 (cont.)

Sensitivity = 144 / (144 + 6)
= 144 / 150

= 0.96
= 96 % sensitive

Specificity = 388 / (388 + 12)
= 388 / 400
= 0.97
= 97 % specific

	Genuinely Positive	Genuinely Negative	Row Total
Test Positive	144	12	156
Test Negative	6	388	394
Column Total	150	400	**550**

Are sensitivity and specificity the same as the positive predictive value (PPV) and negative predictive value (NPV)?

Example 1 (cont.)

Using the values from the example above:

	Genuinely Positive	Genuinely Negative	Row Total
Test Positive	144	12	156
Test Negative	6	388	394
Column Total	150	400	**550**

PPV = 144 / (144 + 12)
= 144 / 156
= 0.923076923… = 92 %

NPV = 388 / (388 + 6)
= 388 / 394
= 0.984771573… = 98 %

So, if a test result is positive, there is a 92 % chance it is correct, if it is negative there is a 98 % chance it is correct.

Example 2

100 people are tested for disease. 15 people have the disease; 85 people are not diseased. So, prevalence is **15%:**

Prevalence of Disease:

$$T_{disease}/T_{total} \times 100,$$

15/100 × 100 = 15%

Truth

Test Result

	Disease (number)	Non Disease (number)	Total (number)
Positive (number)	10 A (True Positive)	40 B (False Positive)	50 $T_{Test\ Positive}$
Negative (number)	5 C (False Negative)	45 D (True Negative)	50 $T_{Test\ Negative}$
	15 $T_{Disease}$	85 $T_{Non\ Disease}$	100 Total

Example 2 (cont.)

Sensitivity is two-thirds, so the test is able to detect two-thirds of the people with disease. The test misses one-third of the people who have disease.

Sensitivity:
A/(A + C) × 100
10/15 × 100 = 67%

The test has 53% specificity. In other words, 45 persons out of 85 persons with negative results are truly negative and 40 individuals test positive for a disease which they do not have.

Specificity:
D/(D + B) × 100
45/85 × 100 = 53%

The sensivity and specificity are characteristics of this test. For a clinician, however, the important fact is among the people who test positive, only 20% actually have the disease.

Example 2 (cont.)

Positive Predictive Value:
$A/(A + B) \times 100$
$10/50 \times 100 = 20\%$

For those that test negative, 90% do not have the disease.

Negative Predictive Value:
$D/(D + C) \times 100$
$45/50 \times 100 = 90\%$

Now, let's change the prevalence..

Homework 1

A sample of people had a test in order to evaluate its effectiveness, the positive and negative test for sick and healthy control people are as follows:

- True positives = 479
- False positive = 74
- True negatives = 226
- False negative = 21

Calculate the following:

- The sensitivity and specificity.
- The positive and negative predictive values.
- The probability of the correct diagnosis.

Test	Sick	Healthy control
Test Positive	479	74
Test Negative	21	226
Total	30	70

Homework 2

100 people are tested for disease. 30 people have the disease; 70 people are not diseased. So, prevalence is **30%:**

Calculate the following:

- The sensitivity and specificity.
- The positive and negative predictive values.
- The probability of the correct diagnosis.

Prevalence of Disease:

$T_{disease}/T_{total} \times 100,$

Truth

Test Result

	Disease (number)	Non Disease (number)	Total (number)
Positive (number)	20 A (True Positive)	33 B (False Positive)	53 $T_{Test\ Positive}$
Negative (number)	10 C (False Negative)	37 D (True Negative)	47 $T_{Test\ Negative}$
	30 $T_{Disease}$	70 $T_{Non\ Disease}$	100 Total

THANKS FOR LISTENING