Epidemiology and Biostatistics

MD

Epidemiology and Biostatistics

Epidemiology

Epidemiology

Learning Objectives

❑❑ Answer questions about epidemiologic measures

❑❑ Use knowledge of screening tests

❑❑ Explain information related to study designs

Epidemiology

EPIDEMIOLOGIC MEASURES

Epidemiology is the study of the distribution and determinants of health-related states within a population. It refers to the patterns of disease and the factors that influence those patterns.

- **Endemic**: the usual, expected rate of disease over time; the disease is maintained without much variation within a region
- **Epidemic**: occurrence of disease in excess of the expected rate; usually presents in a larger geographic span than endemics (*epidemiology* is the study of epidemics)
- **Pandemic**: worldwide epidemic.
- **Epidemic curve**: visual description (commonly histogram) of an epidemic curve is disease cases plotted against time; classic signature of an epidemic is a “spike” in time

Epidemiology

The tools of epidemiology are numbers; the numbers in epidemiology are ratios converted into rates. The denominator is key: who is “at risk” for a particular event or disease state.

To determine the rate, compare the number of actual cases with the number of potential cases:

$$\frac{\text{Actual cases}}{\text{Potential cases}} = \frac{\text{Numerator}}{\text{Denominator}}$$

RATE

Rates are generally, though not always, per 100,000 persons by the Centres for Disease Control (CDC), but can be per any multiplier. (Vital statistics are usually per 1,000 persons.)

Epidemiology

Rates are generally, though not always, per 100,000 persons by the Centres for Disease Control (CDC), but can be per any multiplier. (Vital statistics are usually per 1,000 persons.)

A disease may occur in a country at a regular annual rate, which makes it **endemic**. If there is a sudden rise in the number of cases in a specific month, we say that there is an **epidemic**. As the disease continues to rise and spread to other countries, it becomes a **pandemic**. Thus the terminology is related to both the number of cases and its geographical distribution

Epidemiology

January	February	March	April	May	June	July	August
3	4	3	4	4	4	3	3
5	5	8	8	5	5	5	5

The graph below represents the incidence of 2 diseases (cases in 100,000). Disease 1 is endemic as the rate of disease is consistent month to month with minor variation in the number of cases. Disease 2 experiences an epidemic in March and April in which the number of cases is in excess of what is expected.

Epidemiology

Case:

Consider the following scenario. A Japanese farmer begins to sell meat that is infected with salmonella. Within 2 days, hundreds of villagers begin to experience crampy abdominal pain. This is an example of an epidemic. The sudden rise of Salmonella gastroenteritis in this village is much higher than the average incidence for the given time period.

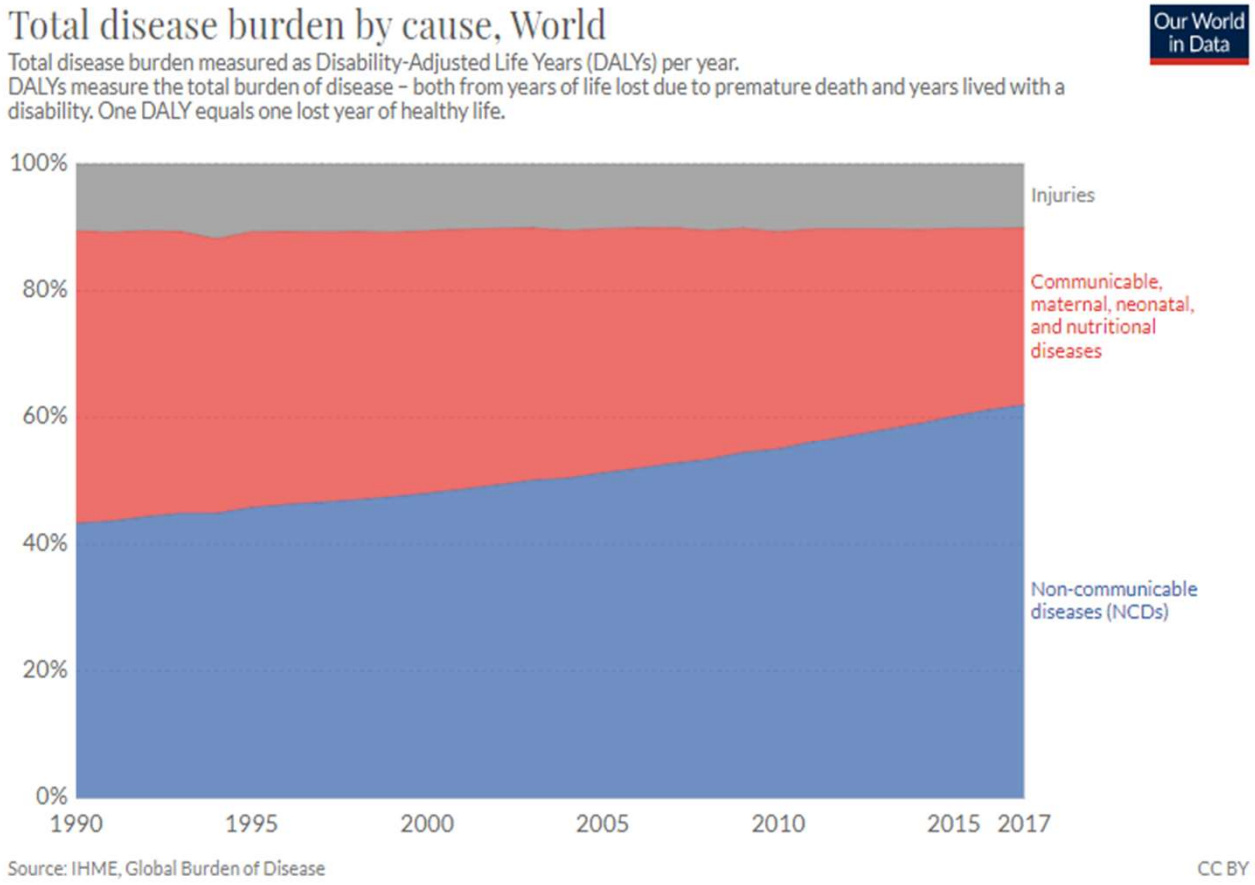
Epidemiology

- Now what if the farmer ships 1,000 pounds of infected beef to other regions of Japan before he realizes what happened? What can one anticipate would happen?

The answer is there would be no change to the endemic rate of gastroenteritis. The farmer is only shipping out 1,000 pounds of beef to a few cities nationwide. Unlike the earlier scenario which addressed the population of a village, this would be the entire nation. If every person who consumes the beef gets gastroenteritis, that number would not significantly increase the national average of cases and would therefore not significantly change the incidence of the disease nationwide.

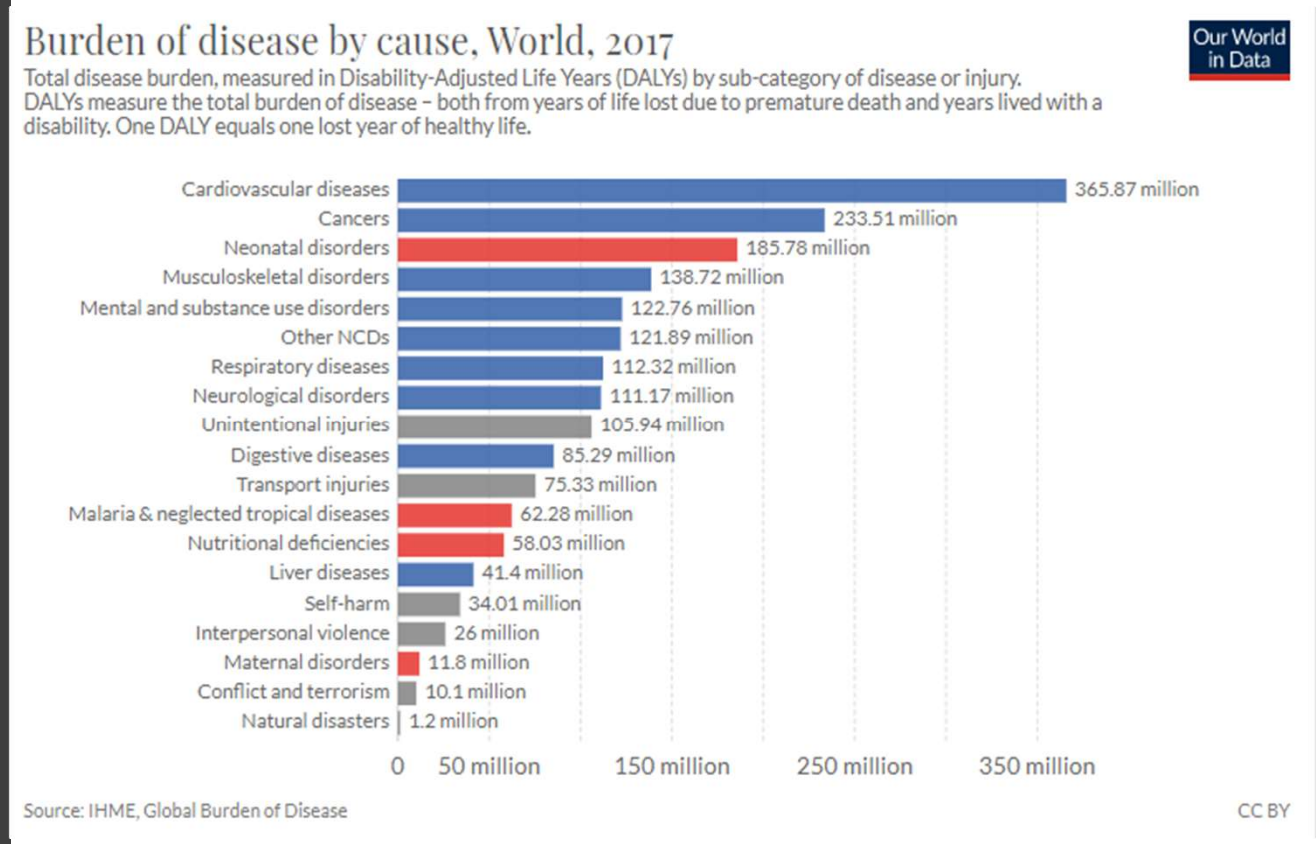
EPIDEMIOLOGY

Global burden of disease



Epidemiology

Burden of disease by cause



Epidemiology

Incidence and Prevalence

Incidence rate (IR) is the rate at which **new events** occur in a population.

- The numerator is the number of **new** events that occur in a defined period.
- The denominator is the population at risk of experiencing this new event during the same period

Epidemiology

Incidence rate

Number of new events in a specified period

Number of persons “exposed to risk” of becoming new cases during this period

- = ×10n

The IR includes only **new cases** of the disease that occurred during the specified period, not cases that were diagnosed earlier. This is especially important when working with infectious diseases such as TB and malaria.

If, over the course of a year, 5 men are diagnosed with prostate cancer, out of a total male study population of 200 (with no prostate cancer at the beginning of the study period), the IR of prostate cancer in this population would be 0.025 (or 2,500 per 100,000 men-years of study).

Epidemiology

Attack Rate is the cumulative incidence of infection in a group of people observed over a period during an epidemic, usually in relation to food-borne illness. It is measured from the beginning of an outbreak to the end of the outbreak.

$$\text{Attack rate} = \frac{\text{Number of exposed people infected with the disease}}{\text{Total number of exposed people}}$$

Attack rate is also called *attack ratio*; consider an outbreak of Norwalk virus in which 18 people in separate households become ill. If the population of the community is 1,000, the overall attack rate is

$18 / 1{,}000 \times 100 = 1.8\%$.

Epidemiology

Prevalence is all persons who experience an event in a population. The numerator is all individuals who have an attribute or disease at a particular point in time (or period of time). The denominator is the population at risk of having the attribute or disease at that point in time or midway through the period.

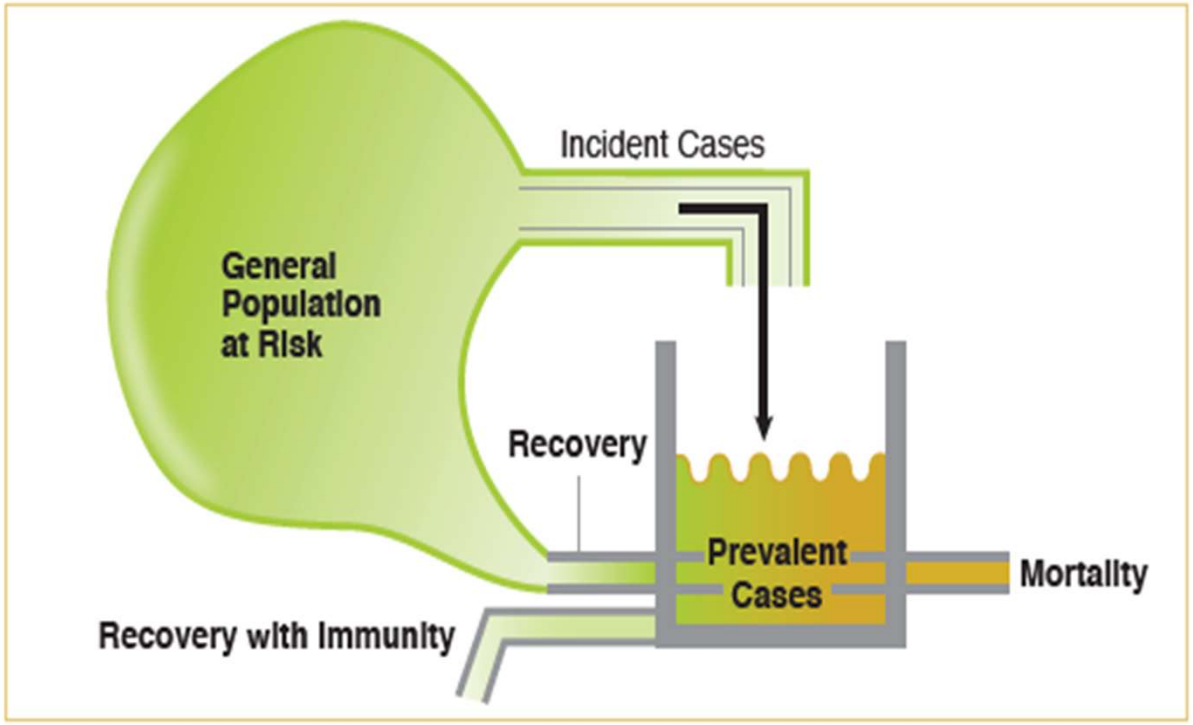
- Prevalence All cases of a disease at a given point / period
- Total population “at risk” for being cases at a given point / period = ×10n

Epidemiology

Point prevalence is useful for comparing disease at different points in time in order to determine whether an outbreak is occurring. We know that the amount of disease present in a population changes over time, but we may need to know how much of a particular disease is present in a population at a single point in time ("snapshot view").

- Perhaps we want to know the prevalence of TB in Community A today. To do that, we need to calculate the point prevalence on a given date. The numerator would include all known TB patients who live in Community A that day. The denominator would be the population of Community A that day.

Epidemiology



- **Period prevalence**, on the other hand, is prevalence during a specified period or span of time. The focus is on *chronic* conditions.
- In the **"prevalence pot**," incident (or new) cases are monitored over time. New cases join pre-existing cases to make up total prevalence.
- Prevalent cases leave the prevalence pot in one of 2 ways: recovery or death.

Epidemiology

A county with a population of 3,000,000 contains 80,000 people who have disease X. There are 200 new cases a year and 40,000 deaths from all causes.

What is the incidence of the disease?

200/ 3,000,000

Epidemiology

Factors effecting incidence and prevalence

New effective treatment is initiated

Incidence ?

Prevalence ?

Epidemiology

New effective vaccine gains widespread use?

Incidence?

Prevalence?

Epidemiology

Number of persons dying from the condition increases?

Incidence?

Prevalence?

Epidemiology

Additional Federal research dollars are targeted to a specific condition?

Incidence

Prevalence?

Epidemiology

Behavioural risk factors are reduced in the population at large?

Incidence?

Prevalence?

Epidemiology

Contacts between infected persons and noninfected persons are reduced

- For airborne infectious disease?
- For noninfectious disease?

Epidemiology

Recovery from the disease is more rapid than it was one year ago?

Incidence?

Prevalence?

Epidemiology

Long-term survival rates for the disease are increasing?

Incidence?

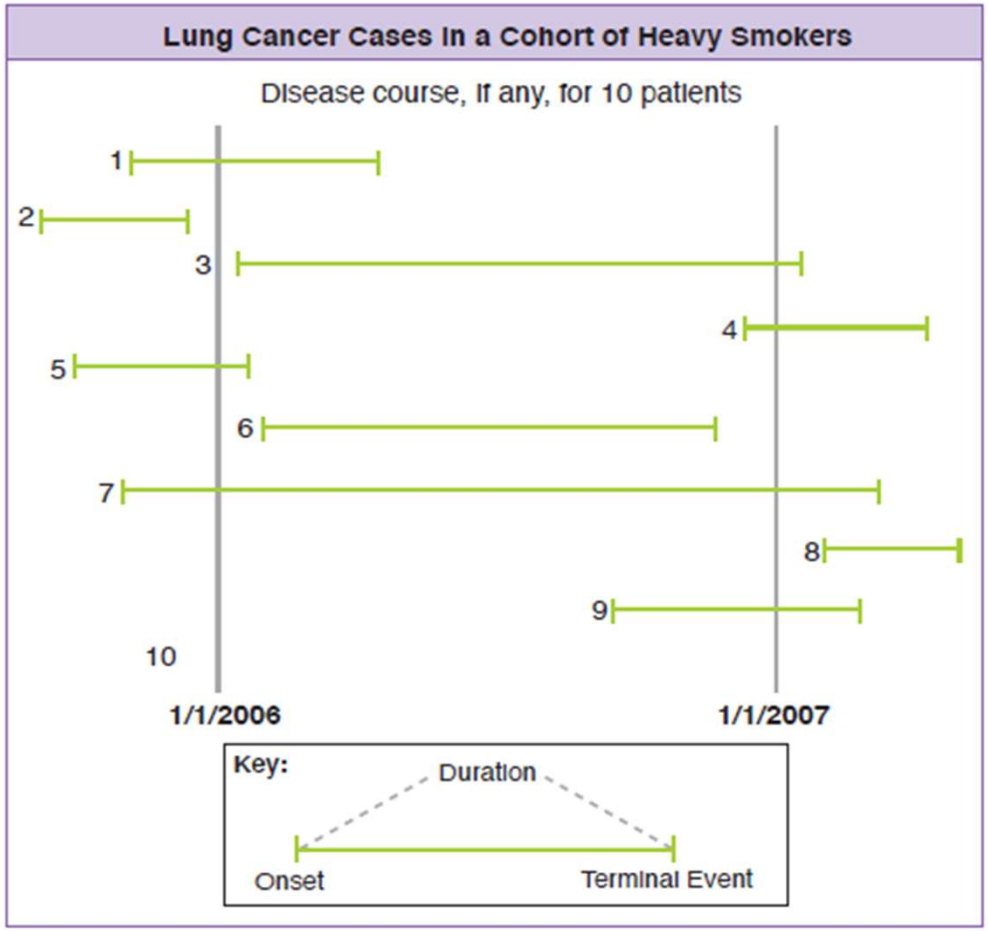
Prevalence?

Epidemiology

What happens if:	Incidence	Prevalence
New effective treatment is initiated	no change	decrease
New effective vaccine gains widespread use	decrease	decrease
Number of persons dying from the condition increases	no change	decrease
Additional Federal research dollars are targeted to a specific condition	no change	no change
Behavioral risk factors are reduced in the population at large	decrease	decrease
Contacts between infected persons and noninfected persons are reduced For airborne infectious disease? For noninfectious disease?	decrease no change	decrease no change
Recovery from the disease is more rapid than it was one year ago	no change	decrease
Long-term survival rates for the disease are increasing	no change	increase

Epidemiology

Incidence and prevalence



Epidemiology

Based on the graph above, **calculate** the following:

- **Prevalence of lung cancer from 1/1/2006–1/1/2007**

Number of patients who "had" lung cancer in this time period from the graph: (7)

Number of patients at risk in this time period: (9) [exclude patient #2 who died before the time period]

Prevalence: (7/9)

Type of prevalence: (period prevalence)

Incidence of lung cancer from 1/1/2006–1/1/2007

Number of patients who developed lung cancer in this time period: (4)

Number of patients at risk in this time period: (6) [exclude patients who were already sick at the start of the time period and those who died before the time period]

Incidence: (4/6)

Epidemiology

Crude, Specific, and Standardized Rates

- **Crude rate** is the actual measured rate for a **whole population**, e.g., rate of myocardial infarction for a whole population. Use caution using the crude rate, though. Imagine that in a given city, there are a lot of older, retired people—the crude rate of myocardial infarction will appear higher even though the rate for each age group has not actually changed.

Epidemiology

- **Specific rate** is the actual measured rate for **subgroup of population**, e.g., “age-specific” or “sex-specific” rate. For instance, the rate of myocardial infarction among people age >65 in the population or the rate of breast cancer among the female population.

- If you are provided specific rates, you can calculate the crude rate. The crude rate of an entire population is a weighted sum of each of the specific rates. The weighted specific rates that are added together is calculated in the table below.

Age specific Mortality Rate

AGE GROUP	NUMBER OF DEATHS	NUMBER OF PEOPLE	DEATH RATE PER 10,000
30-39	400	10,000	400
40-49	600	10,000	600
50-59	800	10,000	800
60-69	1,000	10,000	1,000
70-79	1,200	10,000	1,200
Totals	**4,000**	**50,000**	**800**

Epidemiology

Standardized rate (or adjusted rate) is adjusted to make groups equal on some factor, e.g., age; an “as if ” statistic for comparing groups. The standardized rate adjusts or removes any difference between two populations based on the standardized variable. This allows an “uncontaminated” or unconfounded comparison

Epidemiology

A 73-year-old man comes to the office for follow-up. He was diagnosed with chronic lymphocytic leukemia (CLL) 3 years ago when routine laboratory testing revealed a markedly elevated leukocyte count. He feels well. On examination, he has stable lymphadenopathy. He has been reading about management options should his CLL progress end inquires about an experimental drug that selectively binds malignant lymphocytes. The drug has been shown to significantly prolong survival in patients with stage 3 and 4 CLL, without curing the malignancy. If this new drug were widely used, what changes would be expected in the number of incident and prevalent cases of CLL?

Question

A. Incidence and prevalence decreases
B. Incidence and prevalence increases
C. Incidence decreases prevalence increases
D. Incidence increases and prevalence decreases
E. Prevalence increases and incidence will not change
F. Incidence and prevalence will not change

Epidemiology

Crude Mortality rate

$$\frac{\text{Deaths}}{\text{Population}}$$

Eg

16000 death in certain country.

Or

3 million deaths around the world

Epidemiology

Cause-specific mortality Rate

$$\frac{\text{Deaths from cause}}{\text{Population}}$$

Eg

1000 people died because of LRTI per unit population

Epidemiology

Case-fatality rate

Deaths from cause

Number of persons with the disease/cause

Eg

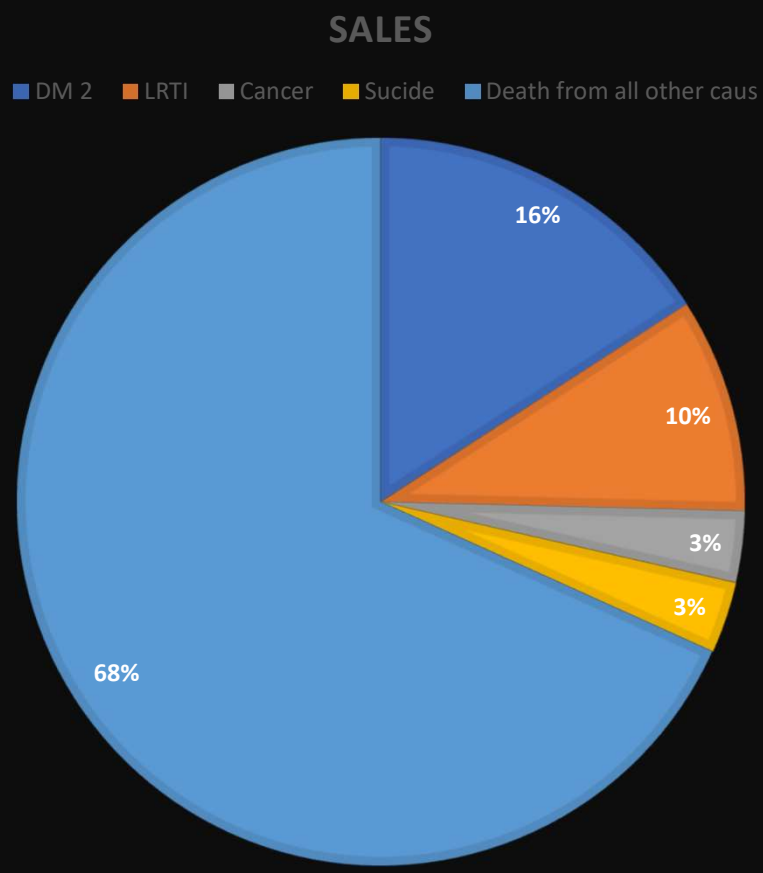
50 people died of cardiovascular disease per unit population

Epidemiology

Proportionate mortality rate (PMR)

Deaths from cause

All deaths



Epidemiology

For example, the city of Hoboken, New Jersey has a population of 50,000. In 2016, the total number of deaths in Hoboken was 400. The number of deaths from lung cancer in Hoboken was 10, while the number of patients with lung cancer diagnosis was 30. Calculate the following:

- Mortality rate in Hoboken for 2016: (400/50,000 × 1,000)
- Cause specific mortality rate for lung cancer in Hoboken for 2016: (10/50,000 × 100,000)
- Case Fatality Rate for lung cancer in Hoboken in 2016: (10/30 × 100)
- PMR for lung cancer in Hoboken in 2016: (10/400 × 100)

Epidemiology

Which of the following will decrease the incidence?

- Multiple Sexual partner – Cervical Cancer
- Obesity - Breast cancer
- Smoking- Lung Cancer
- Using mosquito Nets- Malaria
- PAP smear for Cervical Cancer

Epidemiology

"**Total number of deaths by a particular cause/ Total population**" can be used to calculate

- Crude mortality rate
- Proportional mortality rate
- Cause specific mortality rate
- Standardized mortality rate

Epidemiology

For a given population, below are the details of death by various causes in a year

Tuberculosis - 80

Malaria - 20

Infectious disease- 180

Chronic heart disease 220

Total population- 100,000

What is the **Disease Specific Mortality Rate for Chronic Heart Disease**

A 220/1000

B 22/1000

C 2.2/1000

D 0.22/1000

E 220/500

F 220/100000

Epidemiology

PREVENTION

The goals of prevention in medicine are to **promote health**, **preserve health**, **restore health when it is impaired**, and **minimize suffering and distress**.

These goals aim to minimize both Morbidity and Mortality.

Epidemiology

Primary prevention promotes health at both individual and community levels by facilitating health-enhancing behaviours, preventing the onset of risk behaviours, and diminishing exposure to environmental hazards. **Primary prevention efforts decrease disease incidence**.

Examples include implementation of exercise programs and healthy food programs in schools.

How will it change incidence and prevalence?

Epidemiology

Secondary Prevention screens for risk factors and early detection of asymptomatic or mild disease, permitting timely and effective intervention and curative treatment. **Secondary prevention efforts decrease disease prevalence**.

Examples include recommended annual colonoscopy for patients age >65 and HIV testing for health care workers with needlestick injuries.

How will it decrease incidence and prevalence?

Epidemiology

Tertiary prevention reduces long-term impairments and disabilities and prevents repeated episodes of clinical illness. **Tertiary prevention efforts prevent recurrence and slow progression.**

Examples include physical therapy for spinal injury patients and daily low-dose aspirin for those with previous myocardial infarction.

How will it effect incidence and prevalence?

Epidemiology

Consider a new healthcare bill that is being funded to help wounded war veterans gain access to prosthetic limb replacement. That would be considered tertiary prevention. The patients who would have access to the service have already been injured. The prosthetic devices would help reduce complications of amputation and help their rehabilitation. By improving quality of life and reducing morbidity, that is an implementation of tertiary prevention.

Epidemiology

Now consider a medical student who is asked to wear a nose and mouth mask before entering the room of a patient with meningococcal meningitis.

That would be considered Primary prevention. Because the bacteria in this case can be spread by respiratory contact, the use of the mask will prevent the student from being exposed.

Epidemiology

SCREENING TESTS

Screening tests help physicians to detect the presence of disease, e.g.,

An ELISA test for HIV, the results of which are either positive or negative for disease.

The efficacy of a screening test is assessed by comparing the results to verified sick and healthy populations. For HIV,

We would use a Western blot as a gold-standard.

Epidemiology

The qualifier “true” or “false” is used to describe the correlation between the test results (positive or negative) and the disease (presence or absence).

True-positive (TP): tested positive, actually sick

- In other words, the positive result is true.

False-positive (FP): tested positive, is actually healthy

- In other words, the positive result is false.

True-negative (TN): tested negative, actually healthy

- In other words, the negative result is true.

False-negative (FN): tested negative, is actually sick

- In other words, the negative result is false.

Epidemiology

Screening test

		Disease				
		Present		Absent		Totals
Screening Test Results	Positive	TP	60	FP	70	TP + FP
	Negative	FN	40	TN	30	TN + FN
	Totals	TP + FN		TN + FP		TP + TN + FP + FN

Epidemiology

Measures of Test Performance

Sensitivity and **specificity** are measures of the test performance (and in some cases, physical findings and symptoms). They help to provide additional information in cases where it is not possible to use a gold-standard test and instead a cheaper and easier (yet imperfect) screening test is used. Thin about what would happen if you called the cardiology fellow to do a cardiac catheterization (the gold standard test to diagnose acute myocardial ischemia) on a patient without first having an EKG.

Epidemiology

Sensitivity is the probability of correctly identifying a case of disease. In other words, it is the **proportion of truly diseased persons** in the screened population who are **identified as diseased** by the screening test. This is also known as the **"True Positive rate."**

- Sensitivity = TP/(TP + FN) = true positives/(true positives + false negatives)
- Measures only the distribution of persons with disease
- Uses data from the left column of the 2 × 2 table
- Note: 1-Sensitivity = False Negative Rate

Epidemiology

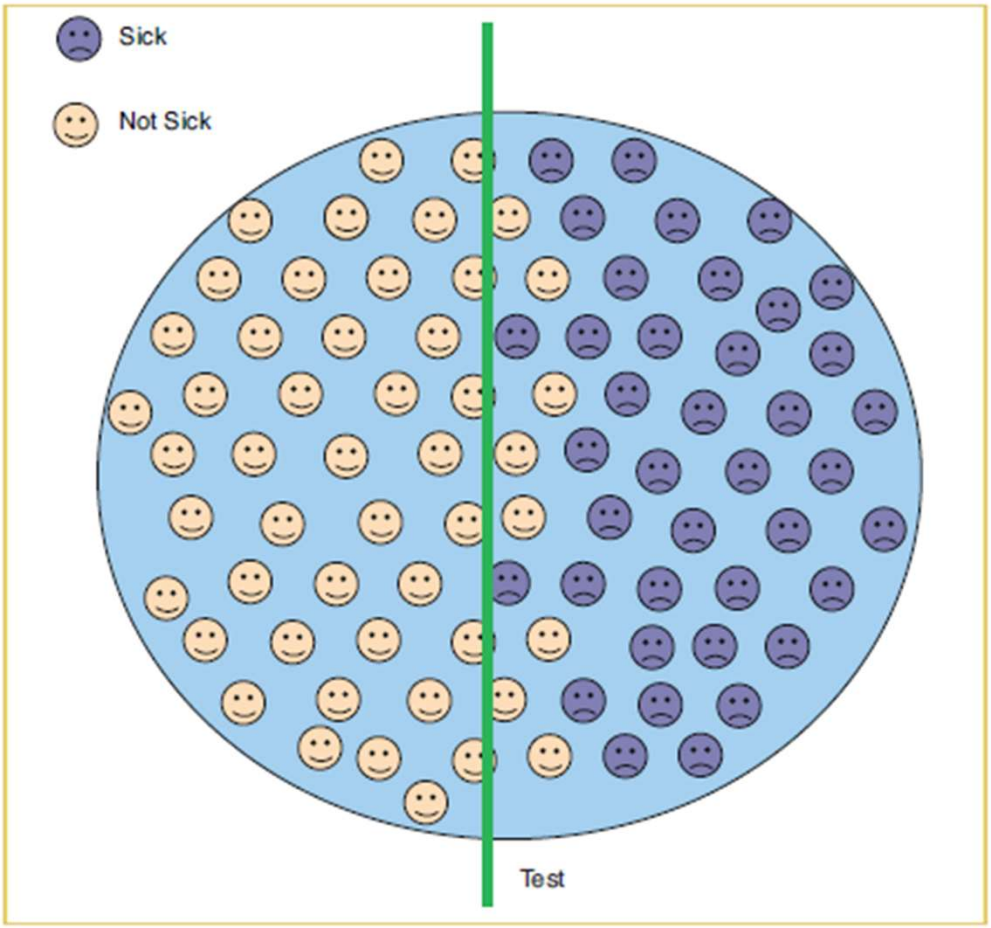
If a test has a **high sensitivity**, then a **negative result** indicates the **absence of the disease.**

For example, temporal arteritis (TA), a large vessel vasculitis that involves branches of the external carotid artery seen in those age >50, always shows elevated ESR. So 100% of patients with TA have elevated ESR. The sensitivity of an abnormal ESR for TA is 100%. If a patient you suspect of having TA has a normal ESR, then the patient does not have TA.

If there are 200 sick people, the sensitivity of a test tells us the capacity of the test to correctly identify these sick people. If a screening test identifies 160 of them as sick (they test positive), then the sensitivity of the test is 160/200 = 80%.

Epidemiology

Note A mnemonic for the clinical use of sensitivity is **SN-N-OUT** (sensitive test-negative-rules out disease).



Epidemiology

Specificity is the probability of correctly identifying disease-free persons.

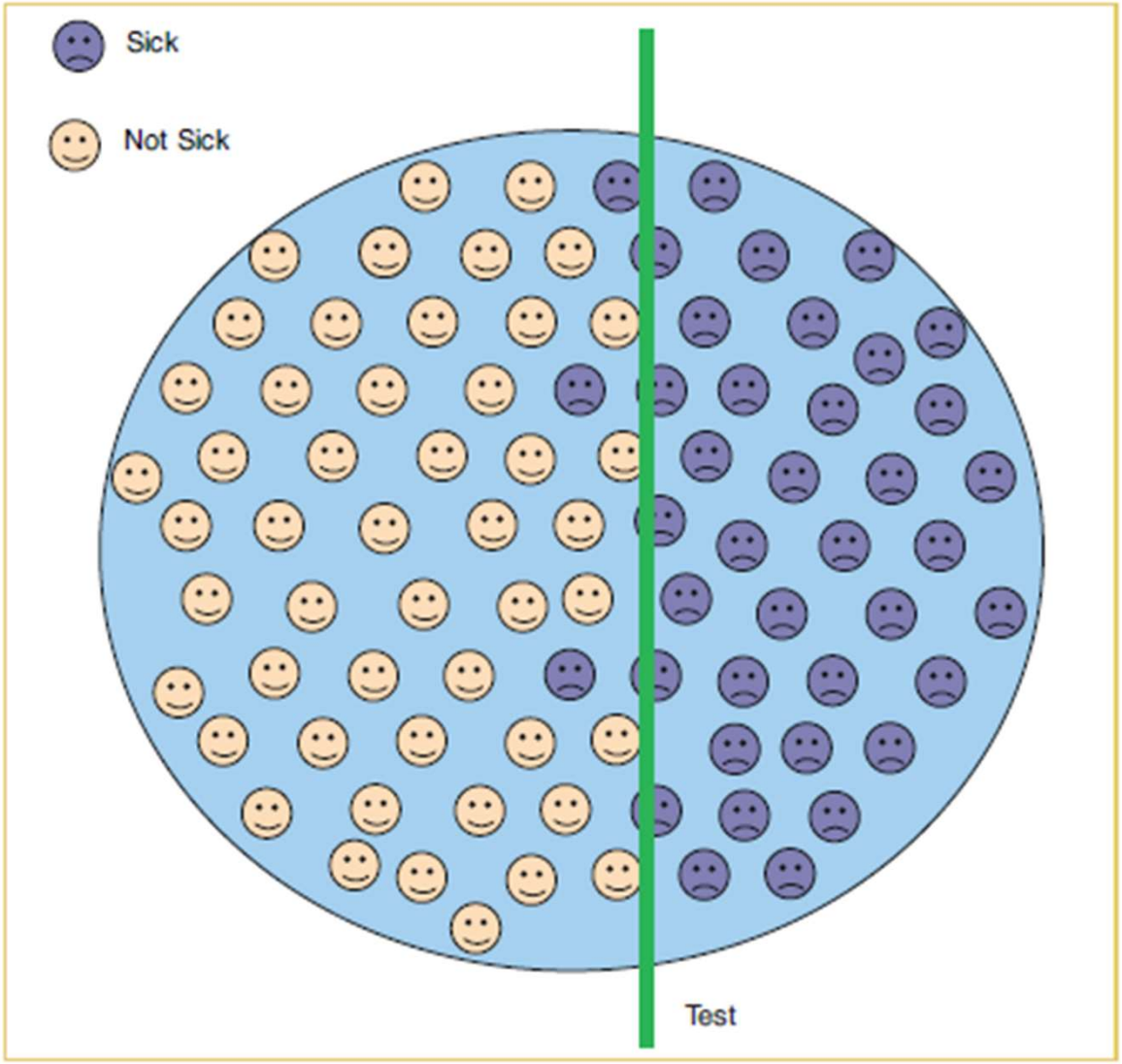
Specificity is the **proportion of truly non diseased persons** who are **identified as non diseased** by the screening test. This is also known as the "true negative rate."

- **Specificity** = TN/(TN + FP) = true negatives/(true negatives + false positives)
- Measures only the distribution of persons who are disease-free
- Uses data from the right column of the 2 × 2 table
- Note: 1-specificity = false positive rate

If a test has a **high specificity**, then a **positive result** indicates the **existence of the disease.** For example, CT angiogram has a very high specificity for pulmonary embolism (97%). A CT scan read as positive for pulmonary embolism is likely true.

Epidemiology

Note A mnemonic for the clinical use of specificity: **SP-I-N** (specific test-positive-rules in disease).



Epidemiology

Post test probability

PPV: Positive predictive value

NPV: Negative predictive value

Epidemiology

Post-Test Probabilities

- **Positive predictive value (PPV)** is the probability of disease in a person who receives a positive test result. The probability that a **person with a positive test is a true positive** (i.e., has the disease) is referred to as the "predictive value of a positive test."

PPV =TP/(TP FP) =True Positives/(True Positives + False Positives)

- PPV measures only the distribution of persons who receive a positive test result.

Epidemiology

Negative predictive value (NPV) is the probability of no disease in a person who receives a negative test result. The probability that a **person with a negative test is a true negative** (i.e., does not have the disease) is referred to as the “predictive value of a negative test.”

NPV= TN/(TN FN) = True negatives/(True negatives + False negatives)

Epidemiology

- NPV measures only the distribution of persons who receive a negative test result.

Accuracy is the total percentage correctly selected, the degree to which a measurement, or an estimate based on measurements, represents the true value of the attribute that is being measured.

Accuracy= (TP+TN)/(TP +TN+ FP+ FN) =(true positives + true negatives)/total screened patients

Epidemiology

Review Questions

Questions 1–3

A screening test identifies 150 out of 1,000 patients to have tuberculosis.

When tested with the gold standard diagnostic test, 200 patients test positive, including 100 of those identified by the screening test.

1. What is the sensitivity of the screening test?
2. What is the specificity of the screening test?
3. What is the positive predictive value?

Epidemiology

Effective Prevalence

Prevalence, which is a quantified measure of disease or cases in the population, is a relevant pre-test probability of disease within the population. The **more disease in a population, i.e., high prevalence**, the greater the probability that a positive test represents actual disease (= greater PPV). The **less disease in a population, i.e., lower prevalence**, the higher the probability that a negative result is true (= greater negative predictive value).

Epidemiology

Consider this example:

Among 80-year-old diabetic patients, the prevalence of kidney failure is higher than in the general population. This increased prevalence makes a physician more likely to believe the results of a screening test that shows kidney failure for an 80-year-old diabetic patient. We intuitively understand that the PPV is higher because this cohort of patients has a higher prevalence of disease.

Epidemiology

Conversely,

If a 15-year-old girl tests positive for a myocardial infarction, a physician will find the results strange and will thus repeat the test to confirm the positive result is not a false-positive. That is because the prevalence of myocardial infarction among teenage girls is so low that a positive result is more likely to be a mistake than a case of an actual myocardial infarction. In a teenage girl, a negative result for myocardial infarction is more likely to be true (high negative predictive value) because there is a very low prevalence of disease in this age group population.

Epidemiology

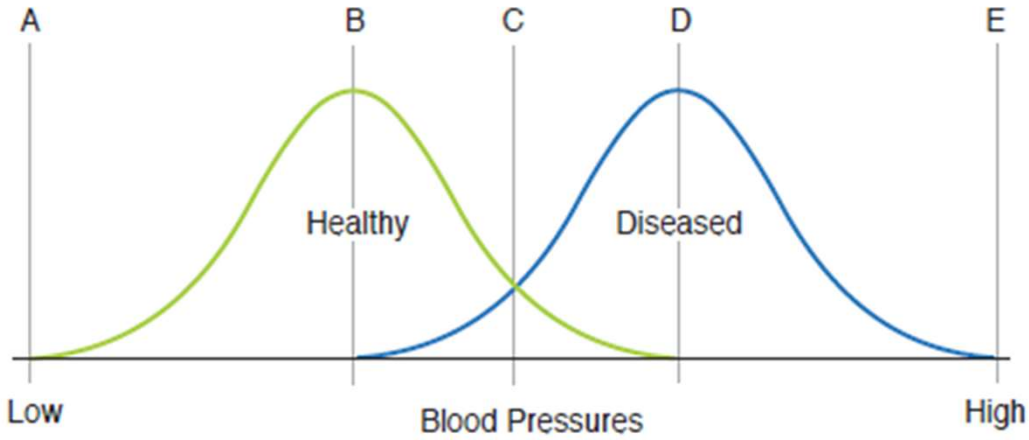
Incidence is a measure of new cases in a population. Increasing the incidence would have no effect on sensitivity or PPV because a screening test can only detect the current presence or absence of disease, not its onset.

Prevalence has no effect on the sensitivity or specificity of a test. Those are metrics of the test and can be changed only by changing the test itself.

Epidemiology

Double Hump Graph

In the graph below, which cut off point provides optimal sensitivity?



Epidemiology

Cut off point B correctly identifies all the sick patients. It has the highest sensitivity (identifies all the sick patients). Cut off D would be the most specific test (it identifies only sick people). Cut off C where the 2 curves intersect is the most accurate. Note, the point of optimum sensitivity equals the point of optimum negative predictive value, while the point of optimum specificity equals the point of optimum positive predictive value.

Consider another example. Which of the following curves indicates the best screening test?

Epidemiology

Epidemiology

Epidemiology

Epidemiology

Epidemiology

Epidemiology

Epidemiology

Epidemiology

Epidemiology