EPIDEMIOLOGY AND BIOSTATISTICS REVIEW, PART II

Danielle Tsingine Chang MSII

Topics to be covered today:

- Incidence vs. prevalence
- Sensitivity
- Specificity
- Negative and Positive Predictive Value
- Case fatality
- Quantifying risk
- T-test
- ANOVA
- Chi-square

Incidence vs. Prevalence

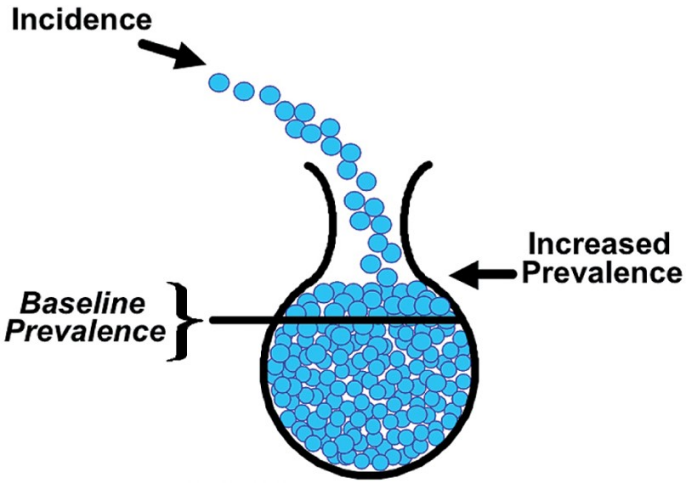
Incidence

- The number of new cases that occur during a specified period of time in a population at risk for developing the disease
- Incidence is a *rate* because it includes time
- = # of new cases of a disease in a specified time period / Population at risk of developing disease during same time period

Prevalence

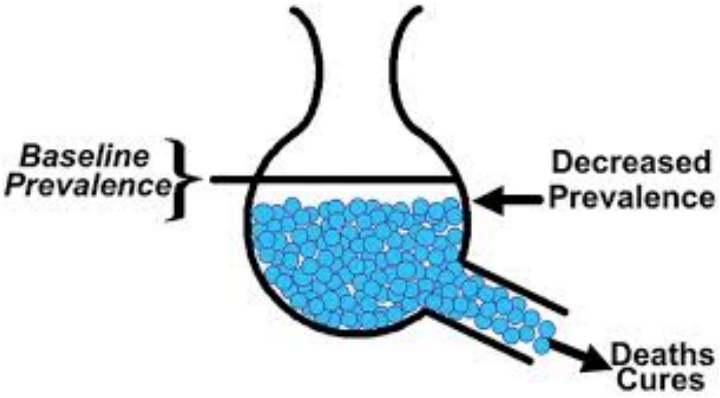
- The number of affected persons present in the population at a specific time divided by the number of persons in the population at that time
- Useful measure of burden of disease in a community
- = # of existing cases / Population at risk
- Prevalence ≈ Incidence x disease duration

Incidence and Prevalence



How can we increase prevalence?

Add new cases i.e. increase the incidence of a disease



How can we decrease prevalence?

Death or cure = reduce the number of diseased persons in the population

Disease X has an incidence of 5 per 1,000 per year and 80% of the people with the disease will die from it. Before 1985, lab test **A** was used to detect the disease. In 1985, a screening test **B** was developed that could detect disease X *two years earlier* than test **A**. However, early detection of the disease did not improve prognosis for disease X. Assume after 1985, all people were screened for disease X using test **B** and that test **B** has a higher sensitivity and specificity than test **A**.

Compare the incidence and prevalence of disease X in 1984 to 1985, the year test **B** was first used.

In 1985, it is true that:

a. Incidence is higher and prevalence is higher than in 1984
b. Incidence is higher in 1984 and prevalence remains the same
c. Incidence is the same and prevalence is higher than in 1984
d. Both incidence and prevalence are the same as in 1984
e. Incidence is the same as in 1984 and prevalence is lower than in 1984

Answer

A. Incidence is higher and prevalence is higher than in 1984

Test **B** catches disease X two years earlier than test **A**, thus the duration of the disease is increased. It is also more sensitive and specific, so you are catching more new cases i.e. increasing the incidence

Prevalence = incidence x duration of disease

The incidence of a chronic disease in a population may be decreased by:

a. Prolonging the lives of persons with the disease
b. Decreasing the case-fatality rate for the disease
c. Improving the treatment of the disease once it has been diagnosed
d. Primary prevention
e. Secondary prevention

Answer

d. Primary prevention

Primary prevention = prevent disease occurrence (i.e. vaccine), thus prevent the numbers of *new cases*

Secondary prevention = early detection of a disease (i.e. screening)

Tertiary prevention = reduce disability from disease (i.e.

Sensitivity and Specificity

- Used more often in a public health setting

- In effect, we are asking, “If we screen a population, what proportion of people who have the disease will be correctly identified?”

Sensitivity

- “How good is the test in correctly identifying those who had the disease?”
- Definition: Proportion of diseased people who were correctly identified as “positive” by the test
- Sensitivity = True Positives/(True Positives + False Negatives) OR = 1 – false-negative rate
- Value close to 100% is desirable for **ruling out** disease
- Used for screening in diseases with **low prevalence**

Specificity

- “How good is the test in correctly identifying those who did not have the disease?”
- Definition: proportion of non-diseased people who are correctly identified as negative by the test
- Specificity = True Negatives / True Negatives + False Positives OR = 1 – false-positive rate
- Value close to 100% is desirable for **ruling in** disease
- Used as a confirmatory test after a positive screening test

Positive and Negative Predictive value

- Used more often in a *clinical* setting

- We are asking, “If the test result is positive in this patient, what is the probability that this patient truly has the disease?”

Positive Predictive Vale

- "What proportion of patients who test positive actually have the disease in question?"
- Definition: proportion of positive test results that are true positive
- PPV = True positives / Total number who tested positive (TP + FP)
- PPV varies directly with prevalence → high prevalence = high PPV

Results of Screening	**Disease**	**No Disease**	Total
Positive	80	100	180
Negative	20	800	820
Total	100	900	1,000

$$\text{PPV} = \frac{\text{TP}}{\text{TP+FP}} = \frac{80}{180} = 44\%$$

Negative Predictive Value

- "If the test is negative, what is the probability that this patient does not have the disease?"
- Definition: Proportion of negative test results that are true negative
- NPV = True Negatives / All people who tested negative (TN + FN)
- NPV varies inversely with prevalence → high prevalence = low NPV

Results of Screening	**Disease**	**No Disease**	Total
Positive	80	100	180
Negative	20	800	820
Total	100	900	1,000

$$\text{NPV} = \frac{\text{TN}}{\text{TN+FN}} = \frac{800}{(20+800)} = 98\%$$

Pulling it all together

A physician examined a population of 1,000 patients in an attempt to detect heart disease. The prevalence of heart disease in this population is known to be 15%. The sensitivity of the physician's exam is 60% and the specificity is 80%. Patients who test positive by the physician are sent for examination by a cardiologist.

1. What is the total number of people who test positive for heart disease based on the physician's exam?

a. 90
b. 260
c. 60
d. 680
e. 740

Answer

1. **b. 260**

Step 1: Prevalence is 15%. If the population is 1,000, then 150 total have heart disease (1,000 X 0.15 = 150)

Step 2: Set up 2x2 table using the given values for sensitivity (60%) and specificity (80%)

Physician's exam results	**Heart Disease**	**No Heart Disease**	Total
Positive	90	170	260
Negative	60	680	740
Total	150	850	1,000

Sensitivity = Proportion of all people with disease who test positive = 60%
So, 150 x 0.6 = 90 → true positives

Specificity = Proportion of all people without disease who test negative = 80%
So, 850 x 0.8 = 680 → true negatives

Thus, total number of people who test positive for heart disease is **260**

A physician examined a population of 1,000 patients in an attempt to detect heart disease. The prevalence of heart disease in this population is known to be 15%. The sensitivity of the physician's exam is 60% and the specificity is 80%. Patients who test positive by the physician are sent for examination by a cardiologist.

2. What is the positive predictive value of the physician's exam?

a. 34.6%
b. 78.8%
c. 85.0%
d. 91.9%

Answer

- 2. a. 34.6%

Physician's exam results	**Heart Disease**	**No Heart Disease**	Total
Positive	90	170	260
Negative	60	680	740
Total	150	850	1,000

PPV = proportion of positive test results that are true positive = TP / (TP + FP)

So, PPV = $\frac{90}{90+170} = \frac{90}{260} = 0.346$ or 34.6%

Quantifying risk

- How do we determine whether a certain disease is associated with a certain exposure?

- To determine whether an association exists, we can use data from case-control and cohort studies

Odds Ratio

- Most often used in case-control studies

- Defined as the ratio of the odds that the cases were exposed to the risk factor to the odds that the controls were exposed

	Cases (with disease)	**Controls (without the disease)**
History of Exposure	A	B
History of No Exposure	C	D

$$\text{OR} = \frac{\frac{A}{C}}{\frac{B}{D}} = \frac{AD}{BC}$$

Relative Risk

- Used most often in cohort studies

- Defined as the risk (i.e. incidence) of developing disease in the exposed group divided by the risk in the unexposed group

	Disease develops	**Disease does not develop**	**Total**	**Incidence rates of disease**
Exposed	A	B	A + B	A / (A + B)
Not exposed	C	D	C + D	C / (C + D)

$$RR = \frac{A/(A+B)}{C/(C+D)}$$

Attributable risk

- Defined as the difference in risk between exposed and unexposed groups, or the proportion of disease occurrences that are attributable to the exposure

$$\text{Attributable risk} = \frac{A}{A+B} - \frac{C}{C+D}$$

Absolute risk reduction (ARR) and Number needed to treat and harm

- ARR is defined as the absolute reduction in risk associated with a treatment as compared to a control

- Number needed to treat is defined as the number of patients who need to be treated for 1 patient to benefit

$$= \frac{1}{Absolute\ risk\ reduction}$$

- Number needed to harm is defined as the number of patients who need to be exposed to a risk factor for 1 patient to be harmed

$$= \frac{1}{Attributable\ risk}$$

A study was performed to determine if an association exists between smoking and lung cancer. In this study, 100 people with a history of smoking tobacco for 10 years and 100 people with no smoking history were followed for 20 years, and the incidence rates for lung cancer were compared in the two groups. The results are below.

	Developed lung cancer	Did not develop lung cancer
(+) Smoking	40	60
(-) Smoking	10	90

1. What is the relative risk for lung cancer in the exposed group?

a. 5

b. 10

c. 4

d. 3

Answer

1. c. 4

RR = Incidence rates in the exposed / incidence rates in the unexposed

RR = (40/40+60) / (10/10+90) = 0.4/0.1 = 4

A study was performed to determine if an association exists between smoking and lung cancer. In this study, 100 people with a history of smoking tobacco for 10 years and 100 people with no smoking history were followed for 20 years, and the incidence rates for lung cancer were compared in the two groups. The results are below.

	Developed lung cancer	Did not develop lung cancer
(+) Smoking	40	60
(-) Smoking	10	90

Calculate the attributable risk and the absolute risk reduction

$$\text{Attributable risk} = \frac{A}{A+B} - \frac{C}{C+D}$$

Or the incidence in the exposed – the incidence in the unexposed

AR = 0.4 – 0.1 = 0.3 → means that 0.3 (or 30%) of people who smoke develop lung cancer as a result of smoking (i.e. the exposer)

Absolute risk reduction percent = (40% of those who smoke develop lung cancer) – (10% of those who do not smoke develop lung cancer) = 30%

→ Describes the difference in risk of developing lung cancer between smokers and nonsmokers.

Mortality rate vs. case-fatality rate

- Mortality rate = a rate calculated by dividing the # of deaths occurring in the population during a stated time period / # of persons at risk of dying during the period
- Can limit the population by age, gender, and disease i.e.
  - annual mortality rate from all causes for children younger than 10 years
  - annual mortality rate from lung cancer in one year
- Key: the denominator represents the entire population at risk of dying from the disease, including those who have the disease and those who *do not (*but are at risk of developing disease)
- Case-fatality rate = # of individuals dying during a specified period of time after disease onset or diagnosis / # of individuals with the specified disease
- Key: denominator is limited to those who already have the disease
- Measure of the severity of the disease

t-test vs. ANOVA vs. Chi-squared

- t-test = checks the difference between the **means of 2 groups**
- ANOVA = checks the difference between the **means of 3 or more groups**
- Chi-square = checks the difference between 2 or more percentages or proportions of categorical outcomes (NOT mean values); used for frequency data rather than for comparison of means

A physician is studying the effects of drug A and drug B on cognitive performance in Alzheimer patients. She administers a memory test to two groups of subjects (those taking drug A and those taking drug B) and compares their mean scores. Which of the following statistical tests would be most appropriate for this purpose?

a) ANOVA
b) Chi-square test
c) Linear regression analysis
d) t-test
e) Multiple linear regression

Answer

- d. t-test

t-test is used to compare two means derived from two samples

Resources

Gordis, Leon. *Epidemiology.* Philadelphia: Saunders Elsevier, 2009.

Le, T. and V. Bhushan. 2013. First aid for the USMLE step 1 2013. New York: McGraw-Hill Medical.

USMLE Step 1 Qbook, Fifth edition