1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

A 28-year-old woman comes to the office to establish care. She recently moved to New York to begin a job as a copy editor at a major newspaper. She has no known medical problems. The patient is in a monogamous relationship with her husband. She usually eats fast food for lunch, and she and her husband cook in the evenings. She goes to the gym about once every 2 weeks. The patient's temperature is 36.7 C (98 F), blood pressure is 118/64 mm Hg, pulse is 60/min, and respirations are 14/min. Her BMI is 24.6 kg/m$^2$. Physical examination is unremarkable. Laboratory results are within the normal range. Counseling this patient regarding diet and exercise habits would be an example of which of the following?

- A. Case finding
- B. Cognitive-behavioral therapy
- C. Community-level intervention
- D. Health promotion
- E. Health risk assessment
- F. Precontemplative stage intervention
- G. Tertiary prevention

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

A 28-year-old woman comes to the office to establish care. She recently moved to New York to begin a job as a copy editor at a major newspaper. She has no known medical problems. The patient is in a monogamous relationship with her husband. She usually eats fast food for lunch, and she and her husband cook in the evenings. She goes to the gym about once every 2 weeks. The patient's temperature is 36.7 C (98 F), blood pressure is 118/64 mm Hg, pulse is 60/min, and respirations are 14/min. Her BMI is 24.6 kg/m$^2$. Physical examination is unremarkable. Laboratory results are within the normal range. Counseling this patient regarding diet and exercise habits would be an example of which of the following?

- A. Case finding [0%]
- B. Cognitive-behavioral therapy [1%]
- C. Community-level intervention [2%]
- ✔ D. Health promotion [80%]
- E. Health risk assessment [6%]
- F. Precontemplative stage intervention [5%]
- G. Tertiary prevention [6%]

**Explanation:**

| Primary, secondary & tertiary prevention | | |
|---|---|---|
| **Level** | **Definition** | **Example(s)** |

**Explanation:**

| Primary, secondary & tertiary prevention | | |
|---|---|---|
| **Level** | **Definition** | **Example(s)** |
| **Primary** | Preventing a disease process from becoming established | • Health promotion (eg, regular exercise, no smoking, weight loss) |
| **Secondary** | Detecting a disease process before it causes symptoms | • Individual case finding (eg, cervical cancer screening)<br>• Community screening (eg, blood pressure screening at state fair) |
| **Tertiary** | Treating a disease to prevent progression/complications | • Disability limitation (eg, blood sugar and blood pressure control in diabetes)<br>• Rehabilitation (eg, physical therapy after stroke) |



According to the World Health Organization, **health promotion** is "the process of enabling people to increase control over their health and its determinants, and thereby improve their health". Examples include improving dietary habits (limiting fat intake, consuming fruits and vegetables), exercising regularly, abstaining from smoking, and losing weight if needed. Health promotion typically falls under **primary**

According to the World Health Organization, **health promotion** is "the process of enabling people to increase control over their health and its determinants, and thereby improve their health". Examples include improving dietary habits (limiting fat intake, consuming fruits and vegetables), exercising regularly, abstaining from smoking, and losing weight if needed. Health promotion typically falls under **primary** prevention, which is preventing a disease process from getting established.

**Secondary prevention** refers to interrupting the disease process before symptoms develop. An example is case finding, which is looking for disease that may be present but **asymptomatic** in patients receiving medical care (eg, age- and gender-appropriate screening) **(Choice A)**. Community screening is an analogous intervention at the community level (eg, measuring hypertension at a state fair). **Tertiary** prevention corresponds to treating an established condition with the goal of minimizing its progression or complications **(Choice G)**.

**(Choice B)** Cognitive-behavioral therapy is a standardized psychotherapeutic modality focusing on identifying and addressing persistent maladaptive thought patterns. It is used as monotherapy or in combination with medication to treat a variety of mental illnesses (eg, anxiety, mood, personality, somatic symptom, eating disorders).

**(Choice C)** Interventions such as imposing taxes on cigarettes or soda or mandating smoke-free establishments can be implemented at the community level to improve the health of the public.

**(Choice E)** Health risk assessments rely on questionnaires that use demographic, medical, lifestyle, and family history information to calculate a patient's "risk age". A patient whose risk age is greater than chronological age has a higher risk of death than the average individual of that chronological age.

**Tertiary** prevention corresponds to treating an established condition with the goal of minimizing its progression or complications **(Choice G)**.

**(Choice B)** Cognitive-behavioral therapy is a standardized psychotherapeutic modality focusing on identifying and addressing persistent maladaptive thought patterns. It is used as monotherapy or in combination with medication to treat a variety of mental illnesses (eg, anxiety, mood, personality, somatic symptom, eating disorders).

**(Choice C)** Interventions such as imposing taxes on cigarettes or soda or mandating smoke-free establishments can be implemented at the community level to improve the health of the public.

**(Choice E)** Health risk assessments rely on questionnaires that use demographic, medical, lifestyle, and family history information to calculate a patient's "risk age". A patient whose risk age is greater than chronological age has a higher risk of death than the average individual of that chronological age.

**(Choice F)** The "stages of change" model (originally developed to help patients quit smoking) assesses an individual's readiness to change a problem behavior. The precontemplative stage is when a patient does not believe there is a problem (eg, smoking) and is unwilling to contemplate a change (eg, quitting). The subsequent stages are contemplation, preparation, action, and maintenance.

**Educational objective:**
Health promotion, according to the World Health Organization, is "the process of enabling people to increase control over their health and its determinants, and thereby improve their health."

Time Spent: 38 seconds



Last updated: [11/3/2015]

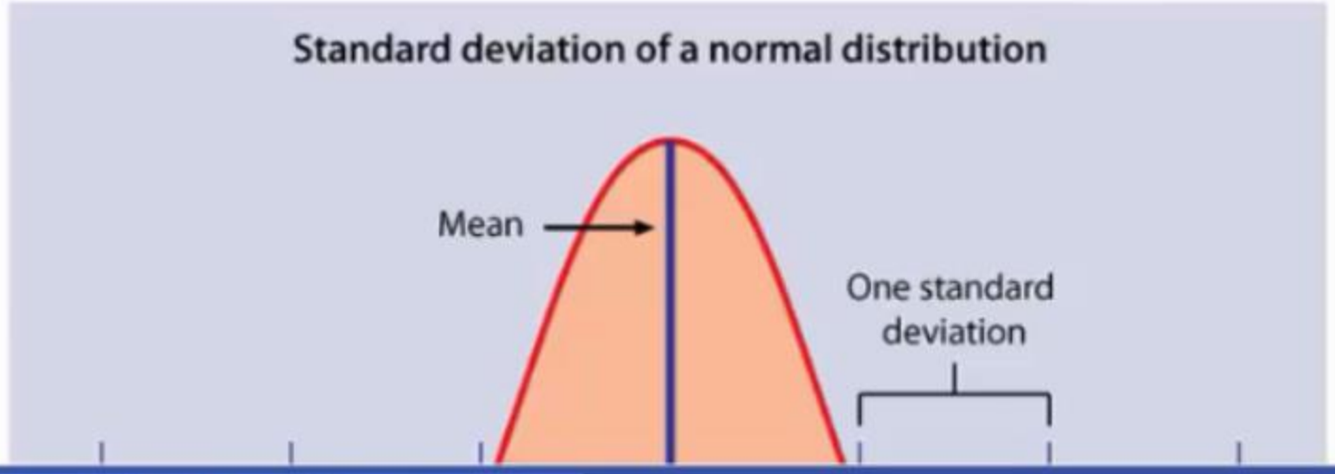
A study was conducted to assess the impact of cholesterol levels on all cause mortality on 400 patients hospitalized with diabetes mellitus-related cardiovascular complications. Serum cholesterol levels were found to be normally distributed among the patients with a mean of 220 mg/dL and standard deviation of 10 mg/dL. Based on the study results, how many patients are expected to have serum cholesterol >240 mg/dL in this study?

- A. 2
- B. 10
- C. 20
- D. 64
- E. 128

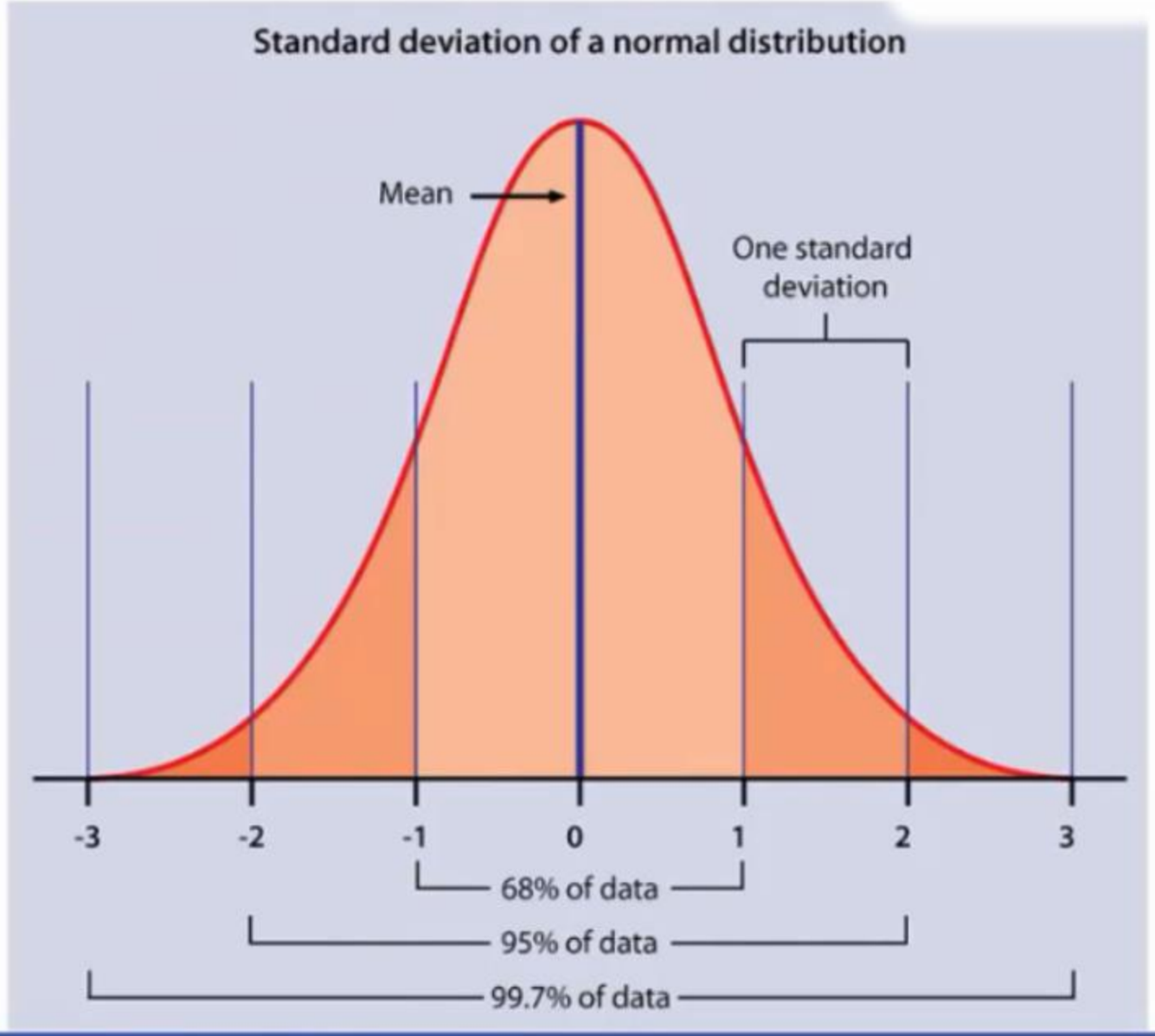
A study was conducted to assess the impact of cholesterol levels on all cause mortality on 400 patients hospitalized with diabetes mellitus-related cardiovascular complications. Serum cholesterol levels were found to be normally distributed among the patients with a mean of 220 mg/dL and standard deviation of 10 mg/dL. Based on the study results, how many patients are expected to have serum cholesterol >240 mg/dL in this study?

- A. 2 [11%]
- ✔ B. 10 [55%]
- C. 20 [23%]
- D. 64 [8%]
- E. 128 [3%]

**Explanation:**



Explanation:



A **normal distribution** refers to a symmetrical, bell-shaped distribution pattern with a fixed proportion of observations lying within a certain distance from the mean. This distance is called the **standard deviation (SD)** and is a measure of the **degree of dispersion** from the mean. When SD is small, data points tend to have minimal variation and are tightly clustered around the mean. In contrast, a large SD implies that the data points are spread out over a large range. The SD is calculated such that **68% of all values lie within 1 SD** from the mean. The remaining 32% of values lie outside of 1 SD, with 16% of these above and 16% below 1 SD from the mean. In addition, **95% of all values are within 2 SDs** from the mean and **99.7% are within 3 SDs**.

With a mean of 220 mg/dL and SD of 10 mg/dL, the given cutoff point of 240 mg/dL is 2 SDs from the mean. Since 95% of values must be within 2 SDs (200-240 mg/dL), 2.5% of values must lie below 200 mg/dL and 2.5% above 240 mg/dL. There are 400 patients in the study, so 10 patients (2.5%) will have a cholesterol level above 240 mg/dL.

**(Choice A)** 2 patients make up 0.5% of the study population. Only 1 or 2 patients at the most would be expected to have cholesterol levels over 250 mg/dL (3 SDs).

**(Choice C)** 20 patients (5% of the study population) would be expected to have cholesterol levels that fall outside of 2 SDs from the mean. However, only half of these will be 2 SDs above the mean (ie, >240 mg/dL); the other half will be 2 SDs below the mean.

**(Choice D)** 64 patients (16% of the study population) will likely have a value that is 1 SD above the mean. However, the cutoff value of 240 mg/dL is 2 SDs above the mean.

are within 3 SDs.

With a mean of 220 mg/dL and SD of 10 mg/dL, the given cutoff point of 240 mg/dL is 2 SDs from the mean. Since 95% of values must be within 2 SDs (200-240 mg/dL), 2.5% of values must lie below 200 mg/dL and 2.5% above 240 mg/dL. There are 400 patients in the study, so 10 patients (2.5%) will have a cholesterol level above 240 mg/dL.

**(Choice A)** 2 patients make up 0.5% of the study population. Only 1 or 2 patients at the most would be expected to have cholesterol levels over 250 mg/dL (3 SDs).

**(Choice C)** 20 patients (5% of the study population) would be expected to have cholesterol levels that fall outside of 2 SDs from the mean. However, only half of these will be 2 SDs above the mean (ie, >240 mg/dL); the other half will be 2 SDs below the mean.

**(Choice D)** 64 patients (16% of the study population) will likely have a value that is 1 SD above the mean. However, the cutoff value of 240 mg/dL is 2 SDs above the mean.

**(Choice E)** 128 patients (32% of the study population) will likely have cholesterol values that lie outside of 1 SD from the mean; however, half of these will be below and half will be above 1 SD from the mean.

**Educational objective:**
In a normal (bell-shaped) distribution:
68% of all values are within 1 standard deviation from the mean
95% of all values are within 2 standard deviations from the mean
99.7% of all values are within 3 standard deviations from the mean

Time Spent: 11 seconds



Last updated: [9/7/2015]

The results of a study investigating a new diagnostic test for acute myocardial infarction (MI) are given in the table below.

| | MI | No MI |
|---|---|---|
| Test positive | 75 | 20 |
| Test negative | 25 | 80 |

What is the sensitivity of the new diagnostic test?

- A. 25%
- B. 37.5%
- C. 50%
- D. 75%
- E. 79%
- F. 80%

The results of a study investigating a new diagnostic test for acute myocardial infarction (MI) are given in the table below.

| | MI | No MI |
|---|---|---|
| Test positive | 75 | 20 |
| Test negative | 25 | 80 |

What is the sensitivity of the new diagnostic test?

- A. 25% [1%]
- B. 37.5% [1%]
- C. 50% [1%]
- ✔ D. 75% [88%]
- E. 79% [6%]
- F. 80% [3%]

**Explanation:**

| | **Positive condition** | **Negative condition** | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |

**Explanation:**

| | Positive condition | Negative condition | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |
| | Sensitivity = TP/(TP+FN) | Specificity = TN/(TN+FP) | |

**FN** = false negative; **FP** = false positive; **TN** = true negative; **TP** = true positive; **NPV** = negative predictive value; **PPV** = positive predictive value.



The data is presented in the standard 2 × 2 (contingency) table format. The **sensitivity** of a test refers to its ability to correctly identify individuals affected with a disease. It is the number of true positives divided by the total number of patients with the disease. Sensitivity should be high in screening tests in order to pick up all cases of a disease (decrease false negatives). Using the generic 2 × 2 (contingency) table:

Sensitivity = TP / (TP + FN), where TP is true positives and FN is false negatives. In this case:

Sensitivity = 75 / (75 + 25) = 75/100 = 0.75 (or 75%)

**(Choice A)** The false negative rate is (1 – sensitivity) = 1 – 0.75 = 0.25 (or

(contingency) table:

Sensitivity = TP / (TP + FN), where TP is true positives and FN is false negatives. In this case:

Sensitivity = 75 / (75 + 25) = 75/100 = 0.75 (or 75%)

**(Choice A)** The false negative rate is (1 – sensitivity) = 1 – 0.75 = 0.25 (or 25%). The false negative rate is not affected by disease prevalence.

**(Choice B)** The value 37.5% is obtained by dividing the TP (75) by the total number of patients in the study (75 + 25 + 20 + 80 = 200).

**(Choice C)** The number of patients with myocardial infarction (MI) based on the gold standard is 100 (= 75 + 25). The total number of patients in the study is 200 (= 75 + 25 + 20 + 80). Therefore, the prevalence of MI in this sample is 100/200 = 0.5 (or 50%).

**(Choice E)** The positive predictive value (PPV) of the test (the probability that a person has the disease given a positive test result) is given by: PPV = TP / (TP + FP), where FP is false positives. In this case, PPV = 75 / (75 + 20) = 75/95 = 0.79 (or 79%).

**(Choice F)** The test specificity (its ability to correctly identify individuals without the disease) is 80 / (80 + 20) = 80/100 = 0.80 (or 80%).

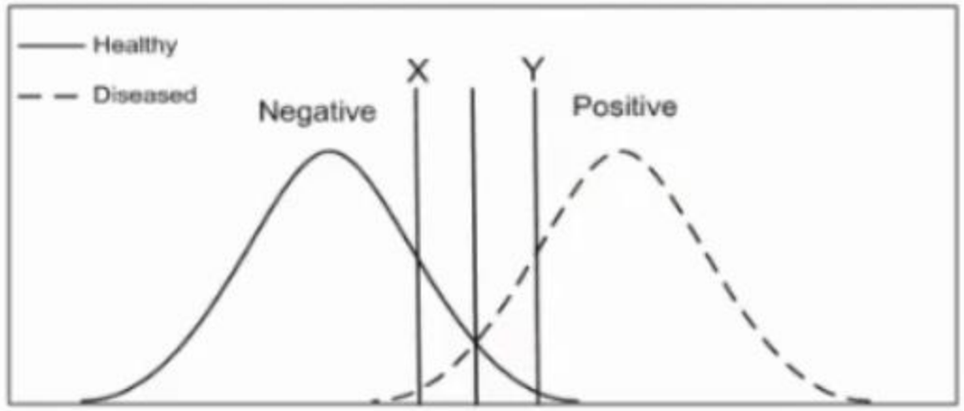
**Educational objective:**
Sensitivity = true positives / (true positives + false negatives). Screening tests should have high sensitivity.

Time Spent: 15 seconds



Last updated: [1/18/2016]

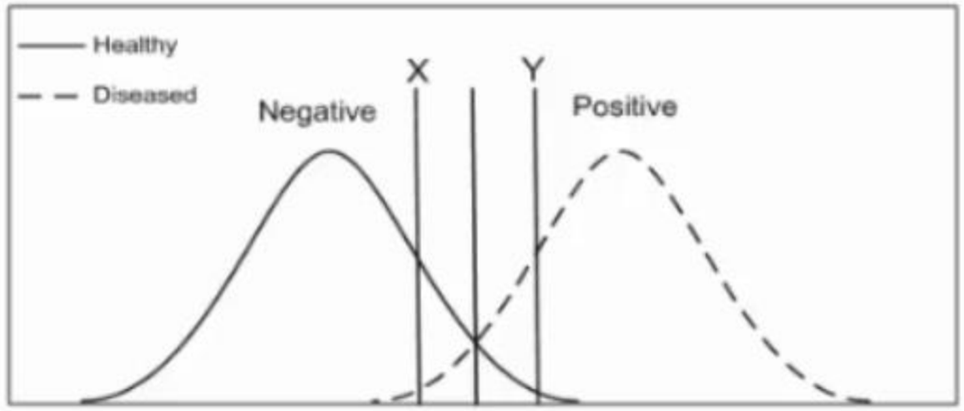
A serum biomarker level is being used for early detection of pancreatic carcinoma. Test results in 200 healthy volunteers and 190 patients with biopsy-proven pancreatic cancer are given in the figure below.




Setting the cutoff value of the serum biomarker level at point X instead of point Y would most likely result in:

- A. Lower sensitivity
- B. Lower number of true positives
- C. Lower number of false positives
- D. Lower positive predictive value
- E. Higher number of false negatives

A serum biomarker level is being used for early detection of pancreatic carcinoma. Test results in 200 healthy volunteers and 190 patients with biopsy-proven pancreatic cancer are given in the figure below.




Setting the cutoff value of the serum biomarker level at point X instead of point Y would most likely result in:

- A. Lower sensitivity [11%]
- B. Lower number of true positives [10%]
- C. Lower number of false positives [19%]
- ✔ D. Lower positive predictive value [43%]
- E. Higher number of false negatives [16%]

**Explanation:**

It is important to first recall the definitions of the parameters of diagnostic tests in

**Explanation:**

It is important to first recall the definitions of the parameters of diagnostic tests in order to correctly answer biostatistics questions.

True positive (TP) - represents the people who have the disease and obtained positive test results.

False negative (FN) - represents the people who have the disease but obtained negative test results (meaning the test failed to identify these people as sick!).

Positive predictive value (PPV) - represents a test's ability to correctly identify those with the disease from all those who had positive results. It is the fraction of those truly with the disease among those with positive results. PPV is dependent on the prevalence of the disease.

Sensitivity - represents the ability of a test to rule out those with the disease. The sensitivity of a test is very important for screening purposes. A test with high sensitivity is one which identifies most patients with the disease. For a high sensitivity test most sick patients will have a positive test result. Sensitivity is not dependent on prevalence.

Specificity - represents the ability of a test to exclude those without the disease. A very specific test is one which has a low false positive rate. In a high specificity test most healthy patients will have a negative test result. Specificity is not dependent on prevalence.

The relationships of these parameters are demonstrated in the equations below:

PPV = TP/ (TP + FP)

Sensitivity = TP/ (TP +FN)

Specificity = TN/ (TN + FP)

prevalence of the disease.

Sensitivity - represents the ability of a test to rule out those with the disease. The sensitivity of a test is very important for screening purposes. A test with high sensitivity is one which identifies most patients with the disease. For a high sensitivity test most sick patients will have a positive test result. Sensitivity is not dependent on prevalence.

Specificity - represents the ability of a test to exclude those without the disease. A very specific test is one which has a low false positive rate. In a high specificity test most healthy patients will have a negative test result. Specificity is not dependent on prevalence.

The relationships of these parameters are demonstrated in the equations below:

PPV = TP/ (TP + FP)
Sensitivity = TP/ (TP +FN)
Specificity = TN/ (TN + FP)

A lower cutoff point for the serum biomarker level will increase the number of positive results and, consequently, increase the sensitivity of the test (**Choice A**). The true positives will also increase (**Choice B**); however, the false positives will have a relatively larger increase (**Choice C**). This results in a decrease in the PPV and the FN (**Choice E**).

**Educational Objective:**
Lowering the cut-off point will increase the sensitivity of a test. The true positives will also increase but the false positives will have a relatively larger increase. This results in a decrease in the PPV and the FN.

Time Spent: 17 seconds



Last updated: [8/19/2015]

A researcher studying physician behavior is interested in how often primary care physicians take the sexual histories of patients during clinic visits. As part of the study, patients who attend a primary care clinic are asked to fill out a questionnaire immediately following a visit with their physician. Once the physicians become aware that their own behavior is being studied, which of the following is most likely to be a potential problem?

- A. Berkson's bias
- B. Hawthorne effect
- C. Lead-time bias
- D. Pygmalion effect
- E. Recall bias

A researcher studying physician behavior is interested in how often primary care physicians take the sexual histories of patients during clinic visits. As part of the study, patients who attend a primary care clinic are asked to fill out a questionnaire immediately following a visit with their physician. Once the physicians become aware that their own behavior is being studied, which of the following is most likely to be a potential problem?

- A. Berkson's bias [9%]
- ✔ B. Hawthorne effect [71%]
- C. Lead-time bias [2%]
- D. Pygmalion effect [15%]
- E. Recall bias [3%]

**Explanation:**

The **Hawthorne effect (observer effect)** is the tendency of study subjects to change their behavior as a result of their awareness that they are being studied. This can impact the observed outcomes, thereby seriously affecting the validity of the study. The Hawthorne effect is commonly seen in studies concerning behavioral outcomes or outcomes that can be influenced by behavioral changes. In this example, physicians (not patients) are the subjects of the study; those physicians who are aware that they are being studied may modify their behavior and start taking sexual histories. To minimize the Hawthorne effect, study subjects can be kept unaware that they are being studied, but this can occasionally pose ethical problems.

**(Choice A)** Berkson's bias refers to selection bias created by choosing hospitalized

1 2 3 4 • 5 • 6 • 7 • 8 • 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

**Explanation:**

The **Hawthorne effect (observer effect)** is the tendency of study subjects to change their behavior as a result of their awareness that they are being studied. This can impact the observed outcomes, thereby seriously affecting the validity of the study. The Hawthorne effect is commonly seen in studies concerning behavioral outcomes or outcomes that can be influenced by behavioral changes. In this example, physicians (not patients) are the subjects of the study; those physicians who are aware that they are being studied may modify their behavior and start taking sexual histories. To minimize the Hawthorne effect, study subjects can be kept unaware that they are being studied, but this can occasionally pose ethical problems.

**(Choice A)** Berkson's bias refers to selection bias created by choosing hospitalized patients as the control group.

**(Choice C)** Lead-time bias refers to the apparent prolongation of survival after applying a screening test that detects a disease earlier than it would have been otherwise detected but without any real effect on prognosis.

**(Choice D)** Pygmalion effect describes the fact that a researcher's beliefs in the efficacy of treatment can potentially affect the outcome. In the classic classroom experiment that first described the Pygmalion effect, a group of students were randomly assigned high intelligence quotient (IQ) scores; their teachers were then told of these artificial results and had higher expectations of this group as a result. The students with the randomly assigned high IQ scores actually performed better, likely because their teachers unconsciously behaved in a manner that would facilitate their success.

**(Choice E)** Recall bias results from inaccurate recall of past exposures by patients.

who are aware that they are being studied may modify their behavior and start taking sexual histories. To minimize the Hawthorne effect, study subjects can be kept unaware that they are being studied, but this can occasionally pose ethical problems.

**(Choice A)** Berkson's bias refers to selection bias created by choosing hospitalized patients as the control group.

**(Choice C)** Lead-time bias refers to the apparent prolongation of survival after applying a screening test that detects a disease earlier than it would have been otherwise detected but without any real effect on prognosis.

**(Choice D)** Pygmalion effect describes the fact that a researcher's beliefs in the efficacy of treatment can potentially affect the outcome. In the classic classroom experiment that first described the Pygmalion effect, a group of students were randomly assigned high intelligence quotient (IQ) scores; their teachers were then told of these artificial results and had higher expectations of this group as a result. The students with the randomly assigned high IQ scores actually performed better, likely because their teachers unconsciously behaved in a manner that would facilitate their success.

**(Choice E)** Recall bias results from inaccurate recall of past exposures by patients. Although recall bias is possible whenever questionnaires are administered, it is unlikely in this case as the patients fill out a form immediately upon leaving the doctor's office.

**Educational objective:**
The Hawthorne effect (observer effect) is the tendency of study subjects to change their behavior as a result of their awareness that they are being studied.

Time Spent: 14 seconds



Last updated: [8/9/2015]

1 2 3 4 5 • 6 • 7 • 8 • 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

A study was conducted to assess the association between simvastatin therapy and serum fibrinogen levels in patients who undergo percutaneous coronary interventions. Serum fibrinogen levels are recorded as either high or normal based on a specific cutoff. The number of patients matching each classification criteria is given in the table below:

| | Fibrinogen high | Fibrinogen normal | |
|---|---|---|---|
| Simvastatin therapy | 43 | 67 | 110 |
| No simvastatin therapy | 32 | 58 | 90 |
| | 75 | 125 | 200 |

Which of the following is the best statistical method for assessing the results of this study?

- A. Two-sample t test
- B. Two-sample z test
- C. Chi-square test
- D. Analysis of variance
- E. Meta-analysis

A study was conducted to assess the association between simvastatin therapy and serum fibrinogen levels in patients who undergo percutaneous coronary interventions. Serum fibrinogen levels are recorded as either high or normal based on a specific cutoff. The number of patients matching each classification criteria is given in the table below:

| | Fibrinogen high | Fibrinogen normal | |
|---|---|---|---|
| Simvastatin therapy | 43 | 67 | 110 |
| No simvastatin therapy | 32 | 58 | 90 |
| | 75 | 125 | 200 |

Which of the following is the best statistical method for assessing the results of this study?

- A. Two-sample t test [32%]
- B. Two-sample z test [3%]
- ✔ C. Chi-square test [57%]
- D. Analysis of variance [6%]
- E. Meta-analysis [2%]

**Explanation:**

1 2 3 4 5 • 6 • 7 • 8 • 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

**Explanation:**

The chi-square test for independence is used to test the association between 2 categorical variables. In this example, patients are divided into 2 groups (simvastatin therapy or no simvastatin therapy) and the number of patients in each group that experience the outcome (high or normal serum fibrinogen levels) is recorded in a 2 x 2 table. Then a chi-square analysis is performed to see if there is a statistical association between simvastatin status and fibrinogen status. If there is a large difference in outcome proportions between the 2 groups, the null hypothesis is rejected and an association between the exposure and the outcome is assumed to be present.

**(Choices A and B)** The two-sample z test and two-sample t test are used to compare 2 group *means*, not categorical variables. The two-sample t test would be appropriate in this example if fibrinogen levels were recorded as numerical values (ie, in mg/dL) instead of categorically.

**(Choice D)** Analysis of variance (ANOVA) is used to compare the *means* of 2 or more groups. For example, an ANOVA analysis could be used to assess for differences in mean blood pressure among 3 sample populations grouped by exercise status (eg, never exercise, exercise occasionally, or exercise frequently).

**(Choice E)** Meta-analysis is an epidemiologic method of pooling the data from several studies to conduct an analysis having a relatively larger statistical power than that of the individual studies.

**Educational objective:**
The chi-square test for independence is used to test the association between 2 categorical variables. In the case of an exposure status and a binomial outcome, patients are divided into 2 groups based on exposure, and the number of patients that experience each outcome is recorded in a 2 x 2 table.

A 54-year-old man with hypertension and hyperlipidemia who came to the emergency department with chest pain wants to know if he is having a heart attack. Test A is newly available for diagnosing myocardial infarction (MI). In a recent study, the results of test A (compared to a gold standard diagnosis of MI) were as follows:

| | MI | No MI |
|---|---|---|
| Test A positive | 200 | 50 |
| Test A negative | 120 | 80 |

The patient has a positive result on test A. Assuming his pre-test probability is equivalent to the prevalence of MI in the study, what is the probability that the patient has an MI?

- A. 40%
- B. 50%
- C. 60%
- D. 80%
- E. 90%

A 54-year-old man with hypertension and hyperlipidemia who came to the emergency department with chest pain wants to know if he is having a heart attack. Test A is newly available for diagnosing myocardial infarction (MI). In a recent study, the results of test A (compared to a gold standard diagnosis of MI) were as follows:

| | MI | No MI |
|---|---|---|
| Test A positive | 200 | 50 |
| Test A negative | 120 | 80 |

The patient has a positive result on test A. Assuming his pre-test probability is equivalent to the prevalence of MI in the study, what is the probability that the patient has an MI?

- A. 40% [3%]
- B. 50% [2%]
- C. 60% [13%]
- ✔ D. 80% [82%]
- E. 90% [1%]

**Explanation:**

The **positive predictive value (PPV)** of a diagnostic test answers the following question: Given a positive test result, what is the probability that a patient has the disease? PPV corresponds to the number of people with the disease who test positive among all those who test positive. Using a standard contingency (2 × 2) table, PPV = a / (a + b)

**Explanation:**

The **positive predictive value (PPV)** of a diagnostic test answers the following question: Given a positive test result, what is the probability that a patient has the disease? PPV corresponds to the number of people with the disease who test positive among all those who test positive. Using a standard contingency (2 × 2) table, PPV = a / (a + b).

**Contingency table example: MI research study**

Generic

| | Disease positive | Disease negative | |
|---|---|---|---|
| Test positive | a<br>True positive | b<br>False positive | a+b |
| Test negative | c<br>False negative | d<br>True negative | c+d |
| | a+c | b+d | |

Test A

| | MI | No MI | |
|---|---|---|---|
| Test A Positive | 200 | 50 | 250 |
| Test A negative | 120 | 80 | 200 |
| | 320 | 130 | |

**MI** = myocardial infarction.



Unlike specificity and sensitivity, PPV varies with disease prevalence. If disease prevalence increases, PPV increases; similarly, PPV decreases with decreasing

| | positive | negative | |
|---|---|---|---|
| Test positive | a True positive | b False positive | a+b |
| Test negative | c False negative | d True negative | c+d |
| | a+c | b+d | |

| | MI | No MI | |
|---|---|---|---|
| Test A Positive | 200 | 50 | 250 |
| Test A negative | 120 | 80 | 200 |
| | 320 | 130 | |

**MI** = myocardial infarction.



Unlike specificity and sensitivity, PPV varies with disease prevalence. If disease prevalence increases, PPV increases; similarly, PPV decreases with decreasing prevalence. In this question, the patient's pre-test probability (which takes into account clinical judgment regarding how likely it is that he has an MI) is assumed to be equivalent to the prevalence of MI in the study, making the results directly translatable. Therefore, PPV = 200 / (200+50) = 200/250 = 0.8 or 80%. In other instances, clinicians may assume that pre-test prevalence equals disease prevalence in the population.

**Educational objective:**
Positive predictive value represents the probability of truly having a disease given a positive test result. It increases with increasing disease prevalence and decreases with decreasing disease prevalence.

Time Spent: 11 seconds



Last updated: [7/29/2015]

1
2
3
4
5
6
7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26
• 27
• 28
• 29
• 30
• 31
• 32
• 33
• 34
• 35
• 36
• 37
• 38
• 39
• 40

A 34-year-old female who recently emigrated from Japan presents to your office with a palpable thyroid node. You proceed with node aspiration that reveals no malignant cells. As you are explaining the test result, the patient asks, "What are the chances that I really do not have cancer?" Which of the following epidemiologic parameters best answers this patient's question?

- A. Sensitivity
- B. Specificity
- C. Positive predictive value
- D. Negative predictive value
- E. Validity

1
2
3
4
5
6
7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26
• 27
• 28
• 29
• 30
• 31
• 32
• 33
• 34
• 35
• 36
• 37
• 38
• 39
• 40

A 34-year-old female who recently emigrated from Japan presents to your office with a palpable thyroid node. You proceed with node aspiration that reveals no malignant cells. As you are explaining the test result, the patient asks, "What are the chances that I really do not have cancer?" Which of the following epidemiologic parameters best answers this patient's question?

- A. Sensitivity [8%]
- B. Specificity [14%]
- C. Positive predictive value [6%]
- ✔ D. Negative predictive value [70%]
- E. Validity [1%]

**Explanation:**

The negative predictive value (NPV) is defined as the probability of being free of a disease if the test result is negative. It is important to remember that the NPV will vary with the pretest probability of a disease. A patient with a high probability of having a disease will have a low NPV if a negative test, whereas a patient with a low probability of having a disease will have a high NPV with a negative test. Specific examples are given below:

1. Thyroid cancer and FNA test results: A patient with a high pre-test probability for having thyroid cancer (e.g., young age, radiation exposure) has a low NPV.
2. HIV and ELISA test results: A patient who belongs to a high-risk group (e.g., multiple sexual partners, admits to not using condoms, IV drug user) has a high pre-test probability; consequently, this patient will have a low NPV. On the

1
2
3
4
5
6
7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26
• 27
• 28
• 29
• 30
• 31
• 32
• 33
• 34
• 35
• 36
• 37
• 38
• 39
• 40

**Explanation:**

The negative predictive value (NPV) is defined as the probability of being free of a disease if the test result is negative. It is important to remember that the NPV will vary with the pretest probability of a disease. A patient with a high probability of having a disease will have a low NPV if a negative test, whereas a patient with a low probability of having a disease will have a high NPV with a negative test. Specific examples are given below:

1. Thyroid cancer and FNA test results: A patient with a high pre-test probability for having thyroid cancer (e.g., young age, radiation exposure) has a low NPV.
2. HIV and ELISA test results: A patient who belongs to a high-risk group (e.g., multiple sexual partners, admits to not using condoms, IV drug user) has a high pre-test probability; consequently, this patient will have a low NPV. On the other hand, a patient who belongs to a low-risk group (e.g., one monogamous sexual partner, consistent condom use, no history of IV drug use) has a low pre-test probability; consequently, this patient will have a high NPV.

*Note: The prevalence of a disease is directly related to the pre-test probability of having the disease and, thus, also affects the NPV.

**(Choices A and B)** The sensitivity and specificity of a test are fixed values which do not vary with the pretest probability of a disease. Most researchers agree that the ideal diagnostic test should have high sensitivity and specificity. In this case, FNA also has a high sensitivity and specificity; however, the statistical parameter being described by the patient and physician was the high negative predictive value.

**(Choice C)** The positive predictive value follows the same concept but applies if the test result is positive.

**(Choice E)** Validity represents the appropriateness of the test (i.e., the test

multiple sexual partners, admits to not using condoms, IV drug user) has a high pre-test probability; consequently, this patient will have a low NPV. On the other hand, a patient who belongs to a low-risk group (e.g., one monogamous sexual partner, consistent condom use, no history of IV drug use) has a low pre-test probability; consequently, this patient will have a high NPV.

*Note: The prevalence of a disease is directly related to the pre-test probability of having the disease and, thus, also affects the NPV.

**(Choices A and B)** The sensitivity and specificity of a test are fixed values which do not vary with the pretest probability of a disease. Most researchers agree that the ideal diagnostic test should have high sensitivity and specificity. In this case, FNA also has a high sensitivity and specificity; however, the statistical parameter being described by the patient and physician was the high negative predictive value.

**(Choice C)** The positive predictive value follows the same concept but applies if the test result is positive.

**(Choice E)** Validity represents the appropriateness of the test (i.e., the test measures what it is supposed to measure). It does not depend on the pretest probability of the disease.

**Educational Objective:**
NPV is the probability of being free of a disease if the test result is negative. Remember that the NPV will vary with the pretest probability of a disease. A patient with a high probability of having a disease will have a low NPV with a negative test but a patient with a low probability of having a disease will have a high NPV with a negative test.

Time Spent: 11 seconds



Last updated: [8/19/2015]

1
2
3
4
5
6
7
8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26
• 27
• 28
• 29
• 30
• 31
• 32
• 33
• 34
• 35
• 36
• 37
• 38
• 39
• 40

Clinicians are developing an enzyme-linked immunosorbent assay test for diagnosing rheumatoid arthritis. The test is designed to detect the presence of serum antibodies against citrullinated proteins. Two test populations with a differing prevalence of rheumatoid arthritis are selected. The clinicians plan to assess the test's performance in the 2 populations by comparing a number of diagnostic test parameters. The difference in disease prevalence is most likely to affect which of the following parameters?

- A. Negative likelihood ratio
- B. Positive likelihood ratio
- C. Positive predictive value
- D. Sensitivity
- E. Specificity

Clinicians are developing an enzyme-linked immunosorbent assay test for diagnosing rheumatoid arthritis. The test is designed to detect the presence of serum antibodies against citrullinated proteins. Two test populations with a differing prevalence of rheumatoid arthritis are selected. The clinicians plan to assess the test's performance in the 2 populations by comparing a number of diagnostic test parameters. The difference in disease prevalence is most likely to affect which of the following parameters?

- A. Negative likelihood ratio [1%]
- B. Positive likelihood ratio [7%]
- ✔ C. Positive predictive value [81%]
- D. Sensitivity [8%]
- E. Specificity [3%]

**Explanation:**

**Common performance measures for diagnostic tests**

| Parameter | Definition | Calculation |
|---|---|---|
| **Sensitivity** | The probability of a diseased person testing positive | $\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$ |

1 2 3 4 5 6 7 8 • 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

**Common performance measures for diagnostic tests**

| Parameter | Definition | Calculation |
|---|---|---|
| **Sensitivity** | The probability of a diseased person testing positive | $\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$ |
| **Specificity** | The probability of a non-diseased person testing negative | $\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}$ |
| **Positive predictive value** | The probability that disease is present given a positive result | $\text{PPV} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$ |
| **Negative predictive value** | The probability that disease is absent given a negative result | $\text{NPV} = \frac{\text{True negatives}}{\text{True negatives} + \text{False negatives}}$ |
| **Positive likelihood ratio** | A ratio representing the likelihood of having the disease given a positive result | $\text{LR+} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$ |
| **Negative likelihood ratio** | A ratio representing the likelihood of having the disease given a negative result | $\text{LR-} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$ |

The positive predictive value (PPV) is the probability that someone who tests positive on a binary diagnostic test actually has the disease. It is calculated by dividing the number of true-positive results by the total number of positive results. If the specificity of the test is <100%, there will be false-positive results (ie, subjects who test positive who do not actually have the disease). Populations having a lower disease prevalence will have fewer true-positive results. Consequently, false positives will make up a greater proportion of all positive test results, and the PPV will therefore be lower. As disease prevalence increases, the PPV will also increase due to a higher proportion of true positives. Similarly, if the sensitivity of the test is <100%, the negative predictive value will increase as the disease prevalence decreases.

**(Choices A and B)** Positive and negative likelihood ratios indicate how a respective positive or negative test result influences the probability of having a disease. Likelihood ratios >1 indicate that the respective test result is associated with the presence of the disease; likelihood ratios <1 mean that the test result is associated with the absence of the disease. Because positive and negative likelihood ratios are based on sensitivity and specificity, they will not be affected by disease prevalence.

**(Choices D and E)** Sensitivity and specificity are not affected by disease prevalence. This is because sensitivity is calculated using true positives and false negatives (only people with the disease), and specificity is calculated using true negatives and false positives (only people without the disease).

**Educational objective:**
Various parameters are used to evaluate the accuracy and usefulness of diagnostic tests. Positive and negative predictive values are influenced by disease prevalence; sensitivity, specificity, and likelihood ratios are not.

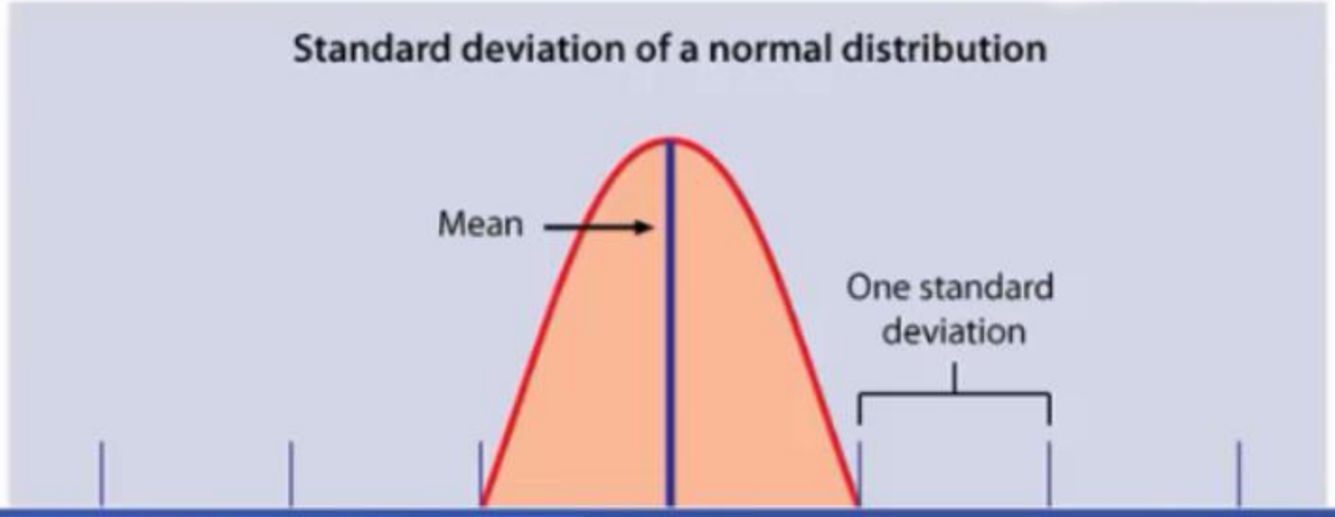
A large study of serum folate levels in a sample of women age 16-45 reveals that this parameter is normally distributed with a mean of 5.0 ng/mL and a standard deviation of 0.5 ng/mL. According to the study results, 95% of serum folate observations in these patients will lie approximately between which of the following limits?

- A. 3.5 and 6.0 ng/mL
- B. 3.5 and 6.5 ng/mL
- C. 4.0 and 6.0 ng/mL
- D. 4.0 and 5.5 ng/mL
- E. 4.5 and 5.5 ng/mL

A large study of serum folate levels in a sample of women age 16-45 reveals that this parameter is normally distributed with a mean of 5.0 ng/mL and a standard deviation of 0.5 ng/mL. According to the study results, 95% of serum folate observations in these patients will lie approximately between which of the following limits?

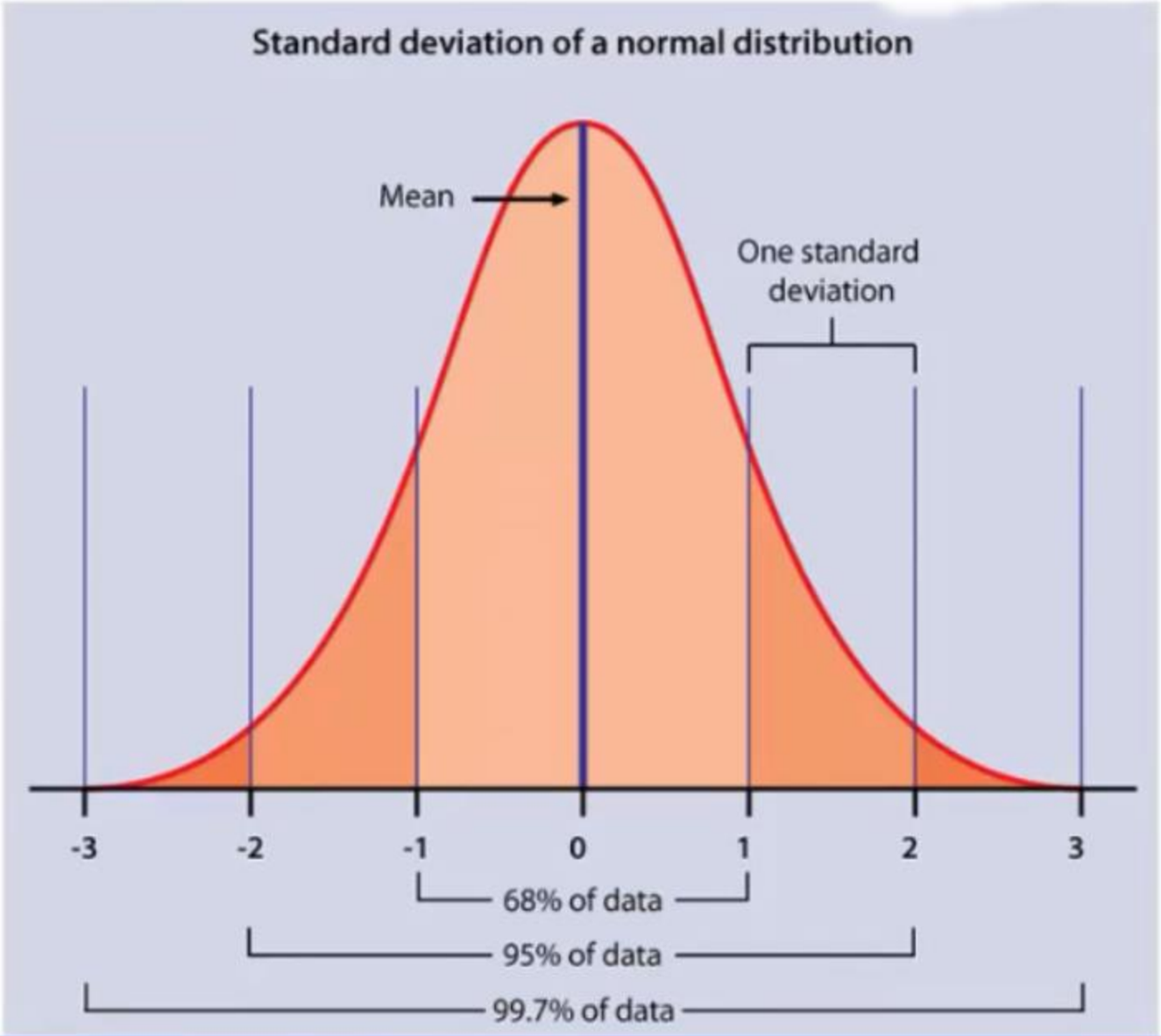
- A. 3.5 and 6.0 ng/mL [1%]
- B. 3.5 and 6.5 ng/mL [5%]
- ✔ C. 4.0 and 6.0 ng/mL [79%]
- D. 4.0 and 5.5 ng/mL [1%]
- E. 4.5 and 5.5 ng/mL [14%]

**Explanation:**



1 2 3 4 5 6 7 8 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

**Explanation:**



A **normal (Gaussian) distribution** is defined as a symmetrical, bell-shaped distribution curve. One of the most important attributes of the normal distribution is the **"68/95/99 rule,"** which states that 68% of all observations lie within 1 standard deviation (SD) of the mean, 95% within 2 SDs of the mean, and 99.7% within 3 SDs of the mean. The SD is a measure of dispersion or variance (how far the measurements are from one another). In this example, the mean folate level is 5.0 ng/mL with a SD of 0.5 ng/mL. Therefore:

- 4.5-5.5 ng/mL is the range within +/- 1 SD of 5.0 ng/mL (given that 5.0 – 1*0.5 = 4.5 ng/mL and 5 + 1*0.5 = 5.5 ng/mL);
- 4.0-6.0 ng/mL is the range within +/- 2 SDs of 5.0 ng/mL (given that 5.0 – 2*0.5 = 4.0 ng/mL and 5.0 + 2*0.5 = 6.0 ng/mL)
- 3.5-6.5 ng/mL is the range within +/- 3 SDs of 5.0 ng/mL (given that 5.0 – 3*0.5 = 3.5 ng/mL and 5.0 + 3*0.5 = 6.5 ng/mL)

Based on the 68/95/99 rule, approximately 95% of observations lie within 2 SDs of the mean, so between 4.0 and 6.0 ng/mL in this case. These are helpful approximations; to be more precise, 95% of the observations lie within 1.96 SDs of the mean and 95.45% of the observations lie within 2 SDs of the mean.

Note that SD is used to describe individual observations in a dataset. A closely related measure called standard error (SE) is used to show how closely sample means are related to population means; SE = SD / $\sqrt{n}$, where n is the sample size. In this example, if SE had been given instead of SD, investigators would be 95% confident that the true mean folate level in the underlying population lies within the mean +/- 2 SE (which is also the range that would include 95% of sample means calculated from repeated samples of the same size taken from that population).

**(Choices A and D)** These limits are asymmetric and are therefore inconsistent with

= 3.5 ng/mL and 5.0 + 3*0.5 = 6.5 ng/mL)

Based on the 68/95/99 rule, approximately 95% of observations lie within 2 SDs of the mean, so between 4.0 and 6.0 ng/mL in this case. These are helpful approximations; to be more precise, 95% of the observations lie within 1.96 SDs of the mean and 95.45% of the observations lie within 2 SDs of the mean.

Note that SD is used to describe individual observations in a dataset. A closely related measure called standard error (SE) is used to show how closely sample means are related to population means; SE = SD / $\sqrt{n}$, where n is the sample size. In this example, if SE had been given instead of SD, investigators would be 95% confident that the true mean folate level in the underlying population lies within the mean +/- 2 SE (which is also the range that would include 95% of sample means calculated from repeated samples of the same size taken from that population).

**(Choices A and D)** These limits are asymmetric and are therefore inconsistent with a normal distribution curve around the mean.

**(Choice B)** Approximately 99.7% of all observations lie within 3 SDs of the mean (3.5-6.5 ng/mL).

**(Choice E)** Approximately 68% of all observations lie within 1 SD of the mean (4.5-5.5 ng/mL).

**Educational objective:**
In a normal (bell-shaped) distribution curve, 68% of observations lie within 1 standard deviation (SD) of the mean, 95% of observations lie within 2 SDs of the mean, and 99.7% of observations lie within 3 SDs of the mean.

Time Spent: 11 seconds



Last updated: [7/29/2015]

A study is performed comparing the effect of tramadol to placebo for painful polyneuropathy. Fifty patients are randomized to 1 of 2 sequences: tramadol followed by placebo or placebo followed by tramadol. Each treatment is delivered for 4 weeks with an interim 1-week washout period. After each treatment period, patients use a 10-point numeric scale to rate pain, paresthesia, and tenderness. Which of the following best describes this study design?

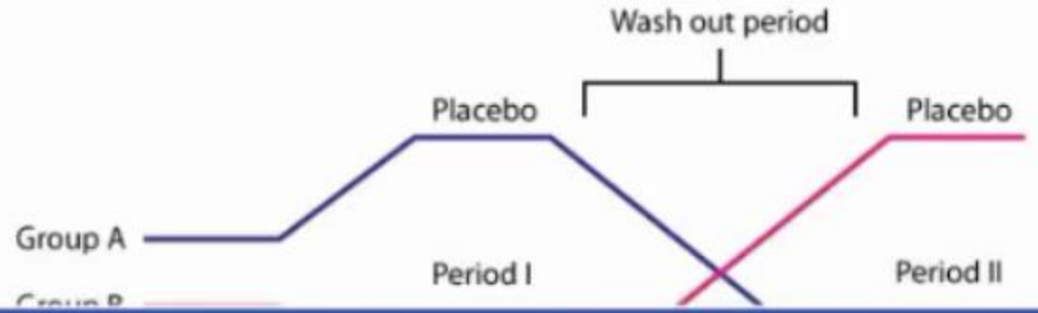
- A. Case-control study
- B. Case series study
- C. Crossover study
- D. Cross-sectional study
- E. Prospective cohort study
- F. Retrospective cohort study

A study is performed comparing the effect of tramadol to placebo for painful polyneuropathy. Fifty patients are randomized to 1 of 2 sequences: tramadol followed by placebo or placebo followed by tramadol. Each treatment is delivered for 4 weeks with an interim 1-week washout period. After each treatment period, patients use a 10-point numeric scale to rate pain, paresthesia, and tenderness. Which of the following best describes this study design?

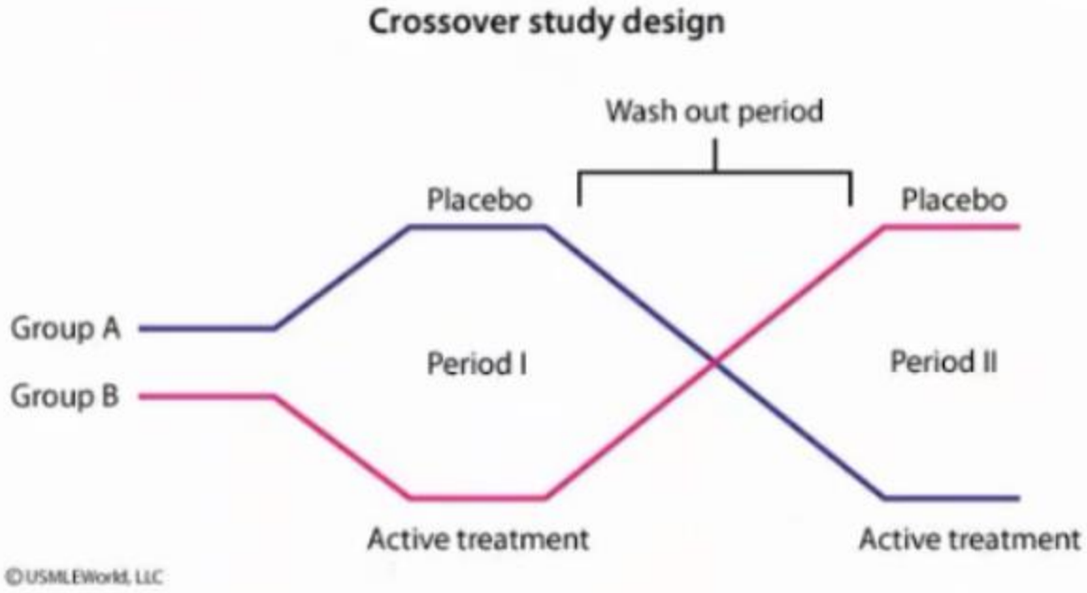
- A. Case-control study [14%]
- B. Case series study [6%]
- ✔ C. Crossover study [56%]
- D. Cross-sectional study [6%]
- E. Prospective cohort study [17%]
- F. Retrospective cohort study [2%]

**Explanation:**

**Crossover study design**



**Explanation:**




In a crossover study, subjects are randomly allocated to a sequence of 2 or more treatments given consecutively. The simplest model is the AB/BA type of study in which subjects allocated to the AB study arm receive treatment A followed by treatment B, and vice versa in the BA arm. Crossover trials allow the patients to serve as their own controls. The principal drawback of crossover trials is that the effects of one treatment may "carry over" and alter the response to subsequent treatments. To limit this disadvantage, a washout (no treatment) period is often added between consecutive treatments. The washout period is designed to be long enough to allow the effects of prior treatment to wear off.

**(Choice A)** A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their

In a crossover study, subjects are randomly allocated to a sequence of 2 or more treatments given consecutively. The simplest model is the AB/BA type of study in which subjects allocated to the AB study arm receive treatment A followed by treatment B, and vice versa in the BA arm. Crossover trials allow the patients to serve as their own controls. The principal drawback of crossover trials is that the effects of one treatment may "carry over" and alter the response to subsequent treatments. To limit this disadvantage, a washout (no treatment) period is often added between consecutive treatments. The washout period is designed to be long enough to allow the effects of prior treatment to wear off.

**(Choice A)** A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

**(Choice B)** A case series is a descriptive study that tracks patients with a known condition (eg, a particular exposure, risk factor, or disease) to document natural history or response to treatment. Unlike a case-control study, a case series is a qualifying study that cannot quantify statistical significance.

**(Choice D)** A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and outcome. It is a snapshot study design that frequently uses surveys. These studies are relatively inexpensive and easy to perform.

**(Choice E & F)** Prospective cohort studies are organized by selecting a group of individuals (ie, cohort), determining their exposure status, and then following them over time for development of the disease of interest. Sometimes the exposure status is determined retrospectively and patients are tracked from the point of exposure onward, typically using medical records.

treatments. To limit this disadvantage, a washout (no treatment) period is often added between consecutive treatments. The washout period is designed to be long enough to allow the effects of prior treatment to wear off.

**(Choice A)** A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

**(Choice B)** A case series is a descriptive study that tracks patients with a known condition (eg, a particular exposure, risk factor, or disease) to document natural history or response to treatment. Unlike a case-control study, a case series is a qualifying study that cannot quantify statistical significance.

**(Choice D)** A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and outcome. It is a snapshot study design that frequently uses surveys. These studies are relatively inexpensive and easy to perform.

**(Choice E & F)** Prospective cohort studies are organized by selecting a group of individuals (ie, cohort), determining their exposure status, and then following them over time for development of the disease of interest. Sometimes the exposure status is determined retrospectively and patients are tracked from the point of exposure onward, typically using medical records.

**Educational objective:**
In a crossover study, subjects are randomly allocated to a sequence of 2 or more treatments given consecutively. A washout (no treatment) period is often added between treatment intervals to limit the confounding effects of prior treatment.

Time Spent: 11 seconds



Last updated: [8/8/2015]

A new test to diagnose urinary tract infections (UTIs) is being evaluated. The sensitivity of the test is 70% and the specificity is 90%. In the study there are 100 patients who actually have UTIs and 200 who actually do not, as determined by the diagnostic gold standard. How many false positives are in this study?

- ○ A. 20
- ○ B. 30
- ○ C. 70
- ○ D. 120
- ○ E. 180

A new test to diagnose urinary tract infections (UTIs) is being evaluated. The sensitivity of the test is 70% and the specificity is 90%. In the study there are 100 patients who actually have UTIs and 200 who actually do not, as determined by the diagnostic gold standard. How many false positives are in this study?

- ✔ A. 20 [70%]
- B. 30 [18%]
- C. 70 [6%]
- D. 120 [3%]
- E. 180 [5%]

**Explanation:**

False positives describe people who test positive for a disease but do not actually have it. In the table below they are represented by square b.

| | Disease positive | Disease negative | |
|---|---|---|---|
| **Test positive** | a<br>True positive | b<br>False positive | a+b |
| **Test negative** | c<br>False negative | d<br>True negative | c+d |
| | a + c | b + d | |

**Explanation:**

False positives describe people who test positive for a disease but do not actually have it. In the table below they are represented by square b.

| | Disease positive | Disease negative | |
|---|---|---|---|
| **Test positive** | a<br>True positive | b<br>False positive | a+b |
| **Test negative** | c<br>False negative | d<br>True negative | c+d |
| | a + c | b + d | |

**Sensitivity = a/(a+c)**
**Specificity = d/(b+d)**



The number of false positives can be calculated using the test's specificity and total number of patients without UTIs. The specificity of a test refers to the number of true negatives divided by the total number of patients without the disease (d / [b + d]). In this case, the specificity is 90% and the total number of patients without UTIs (b + d) equals 200. Using this information, the number of true negatives (d) can be calculated:

Specificity = d / (b + d)
0.9 = d / 200
d = 180

Specificity = d / (b + d)
0.9 = d / 200
d = 180

Since the total number of patients without the disease (b + d) equals 200 and the number of true negatives (d) equals 180, the number of false positives (b) is 20.

For the sake of thoroughness, the two-by-two table can be completed as follows. Sensitivity refers to the number of true positives divided by all people with the disease (a / [a + c]). Repeating the same exercise as above gives the completed table:

Sensitivity = a / (a + c)
0.7 = a / 100
a = 70
c = 100 – 70 = 30

| | + UTI | - UTI | |
|---|---|---|---|
| Test positive | 70 | 20 | 90 |
| Test negative | 30 | 180 | 210 |
| | 100 | 200 | |

The number of true negatives and false positives can be more quickly calculated using the following equations:

True negatives = (Specificity) * (Number of patients actually without the disease)
False positives = (1 - Specificity) * (Number of patients actually without the

completed table:

Sensitivity = a / (a + c)
0.7 = a / 100
a = 70
c = 100 – 70 = 30

| | + UTI | - UTI | |
|---|---|---|---|
| Test positive | 70 | 20 | 90 |
| Test negative | 30 | 180 | 210 |
| | 100 | 200 | |

The number of true negatives and false positives can be more quickly calculated using the following equations:

True negatives = (Specificity) * (Number of patients actually without the disease)
False positives = (1 - Specificity) * (Number of patients actually without the disease)

**Educational objective:**
Specificity is the number of true negatives divided by the total number of subjects actually without the disease.

True negatives = (Specificity) * (Number of patients actually without the disease)
False positives = (1 - Specificity) * (Number of patients actually without the disease)

Time Spent: 12 seconds



Last updated: [10/7/2015]

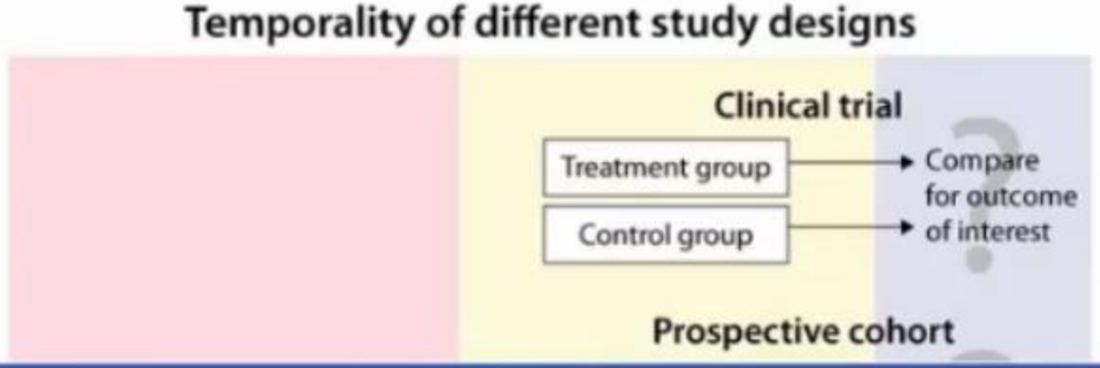
Researchers are studying the relationship between essential hypertension and a common mutation in the structure of a sodium channel protein. A study population is randomly selected and blood samples are obtained for leukocyte genotyping. The prevalence of hypertension is determined based on mean blood pressure measurements obtained using standardized ambulatory blood pressure monitoring conducted over 1 week. Based on the analysis results, the researchers conclude that the sodium channel structure mutation is associated with hypertension. Which of the following best describes the study design used by the investigators?

- A. Case-control study
- B. Cross-sectional study
- C. Prospective cohort study
- D. Randomized clinical trial
- E. Retrospective cohort study

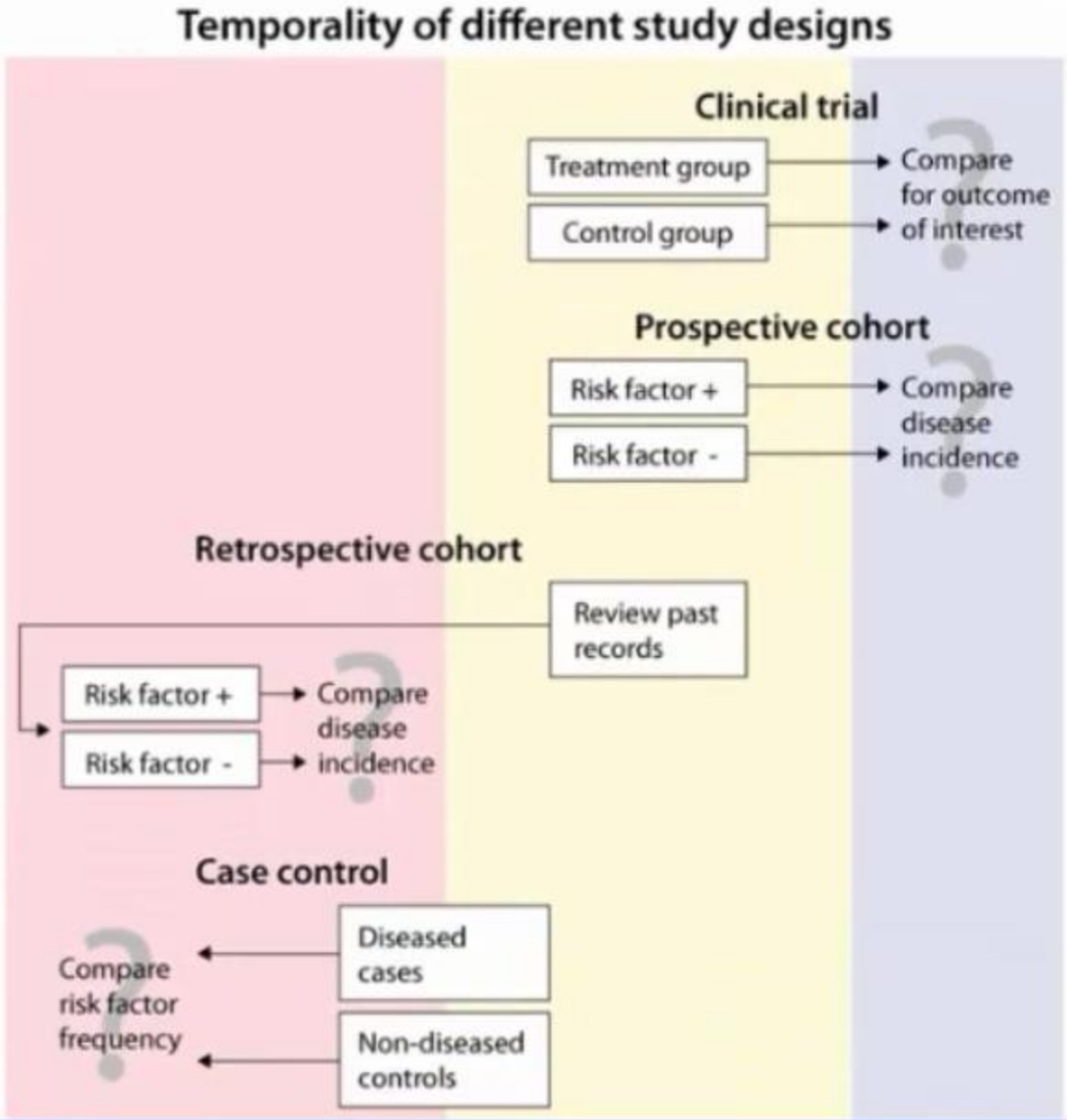
Researchers are studying the relationship between essential hypertension and a common mutation in the structure of a sodium channel protein. A study population is randomly selected and blood samples are obtained for leukocyte genotyping. The prevalence of hypertension is determined based on mean blood pressure measurements obtained using standardized ambulatory blood pressure monitoring conducted over 1 week. Based on the analysis results, the researchers conclude that the sodium channel structure mutation is associated with hypertension. Which of the following best describes the study design used by the investigators?

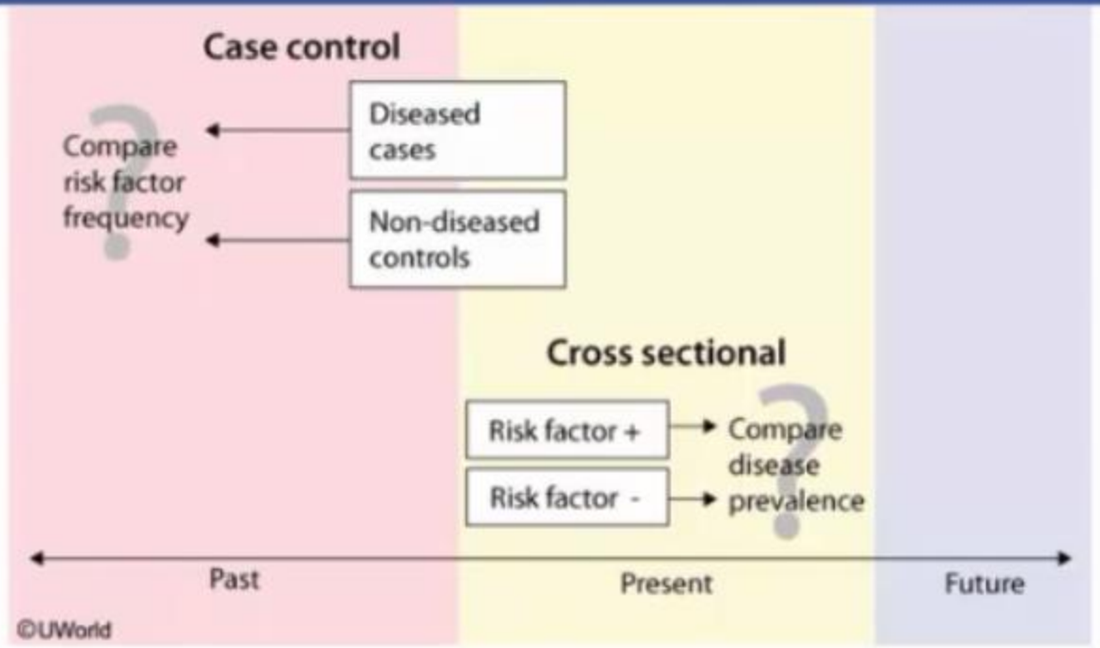
- A. Case-control study [11%]
- ✔ B. Cross-sectional study [56%]
- C. Prospective cohort study [17%]
- D. Randomized clinical trial [10%]
- E. Retrospective cohort study [6%]

**Explanation:**



**Explanation:**






A **cross-sectional** study (also known as a prevalence study) simultaneously measures exposures and outcomes. The cross-sectional study has a "**snapshot**" design that is frequently used in surveys, mostly because it is inexpensive and easy to perform. In this example, a snapshot was obtained of individuals randomly selected from the population; their blood samples were analyzed for the presence of the sodium channel protein mutation and the prevalence of hypertension was calculated. The subjects' blood pressure was measured over 7 days to obtain an average measurement (likely to limit the results being impacted by white-coat hypertension and other transient causes of hypertension). The major limitation of a cross-sectional study design is that a temporal relationship between exposure and outcome is not always clear. However, in this case, demonstrating a temporal relationship was straightforward because the possession of a specific genotype clearly precedes hypertension.

**(Choice A)** A case-control study is designed by selecting individuals with a particular disease (cases), individuals without that disease (controls), and then evaluating previous exposure status. This study would have been classified as a case-control study if it had explicitly recruited individuals with (cases) and without (controls) hypertension (rather than randomly selecting a sample from a population) and evaluated their sodium channel mutation patterns.

**(Choices C and E)** A prospective cohort study would have taken individuals without hypertension from the population, analyzed their blood samples to determine the distribution of the sodium channel mutation, and followed them over time (years) to determine the proportion of new cases of hypertension in patients with and without the mutation. In this example, although hypertension was measured over several days, the objective was not to document incidence of new cases of hypertension (which would be unlikely to develop over 1 week) but to measure the prevalence of hypertension while ensuring that average blood pressure measurements were obtained.

A retrospective design would have also assessed incidence (not prevalence) of hypertension compared to an earlier period of time, based on a chart review of historic data. Prospective and retrospective cohort studies are organized by selecting a group of individuals (ie, a cohort) who do **not** have the disease of interest (eg, hypertension), determining their exposure status, and then following them (forward in time or from a point in the past to the present) to assess for the development of the disease.

**(Choice D)** A randomized clinical trial directly compares ≥2 treatments. Usually, the subjects are randomly assigned to experience a specific exposure (eg, a medication) or no exposure (eg, placebo) and are then followed to assess for the outcome of interest (eg, disease).

**(Choices C and E)** A prospective cohort study would have taken individuals without hypertension from the population, analyzed their blood samples to determine the distribution of the sodium channel mutation, and followed them over time (years) to determine the proportion of new cases of hypertension in patients with and without the mutation. In this example, although hypertension was measured over several days, the objective was not to document incidence of new cases of hypertension (which would be unlikely to develop over 1 week) but to measure the prevalence of hypertension while ensuring that average blood pressure measurements were obtained.

A retrospective design would have also assessed incidence (not prevalence) of hypertension compared to an earlier period of time, based on a chart review of historic data. Prospective and retrospective cohort studies are organized by selecting a group of individuals (ie, a cohort) who do **not** have the disease of interest (eg, hypertension), determining their exposure status, and then following them (forward in time or from a point in the past to the present) to assess for the development of the disease.

**(Choice D)** A randomized clinical trial directly compares ≥2 treatments. Usually, the subjects are randomly assigned to experience a specific exposure (eg, a medication) or no exposure (eg, placebo) and are then followed to assess for the outcome of interest (eg, disease).

**Educational objective:**
In a cross-sectional study, exposure and outcome are measured simultaneously at a particular point in time ("snapshot study"). In other study designs, a certain time period separates the exposure from the outcome.

Time Spent: 15 seconds



Last updated: [11/24/2015]

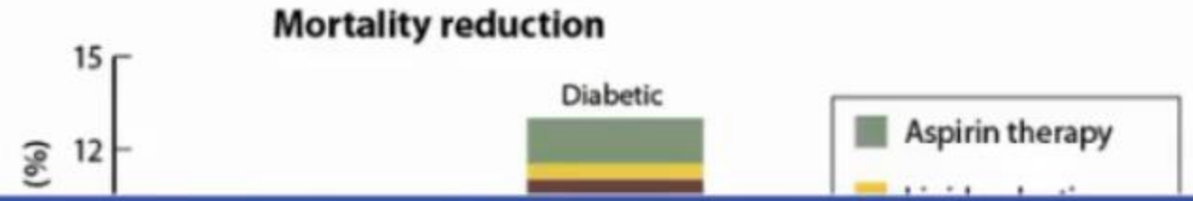
A 50-year-old man comes to the physician for a routine check-up. He has a 20-pack-year smoking history and consumes alcohol occasionally. He eats a balanced diet and consumes fried foods and red meat several times a week. He is not very active and spends much of his free time watching television and sitting at his computer. His family history is significant for hypertension in his mother and bladder cancer in his father. His blood pressure is 160/90 mm Hg and heart rate is 70/min. Laboratory testing is significant for a fasting blood glucose level of 155 mg/dL. Which of the following interventions is likely to have the greatest effect on reducing this patient's mortality risk?

- ○ A. Annual fecal occult blood test
- ○ B. Annual prostate-specific antigen test
- ○ C. Daily aspirin
- ○ D. Dietary modification
- ○ E. Regular exercise
- ○ F. Smoking cessation
- ○ G. Tight glycemic control

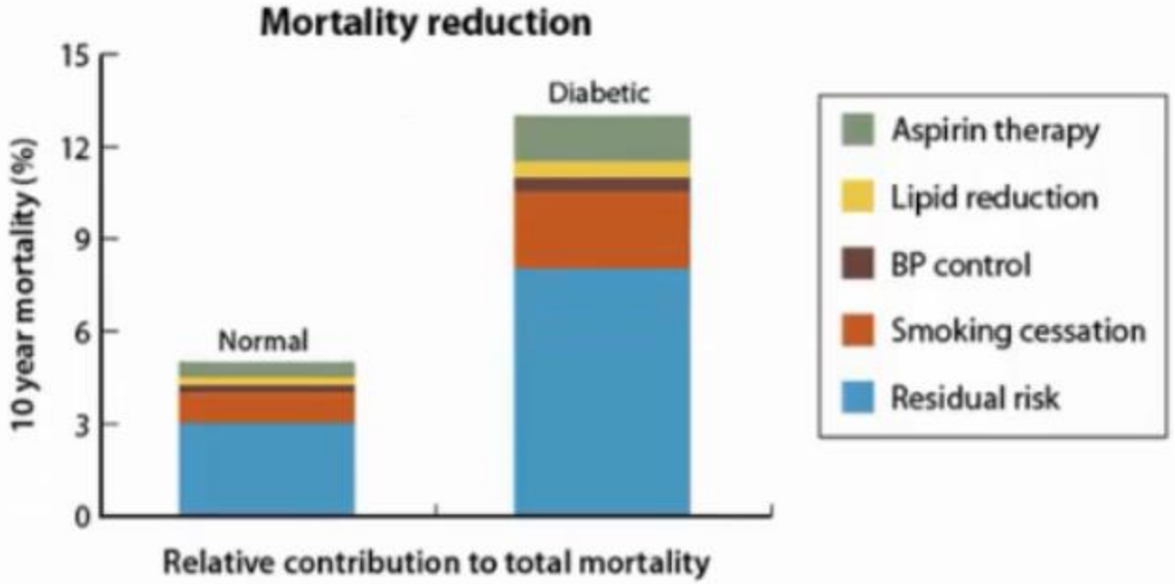
A 50-year-old man comes to the physician for a routine check-up. He has a 20-pack-year smoking history and consumes alcohol occasionally. He eats a balanced diet and consumes fried foods and red meat several times a week. He is not very active and spends much of his free time watching television and sitting at his computer. His family history is significant for hypertension in his mother and bladder cancer in his father. His blood pressure is 160/90 mm Hg and heart rate is 70/min. Laboratory testing is significant for a fasting blood glucose level of 155 mg/dL. Which of the following interventions is likely to have the greatest effect on reducing this patient's mortality risk?

- A. Annual fecal occult blood test [0%]
- B. Annual prostate-specific antigen test [0%]
- C. Daily aspirin [4%]
- D. Dietary modification [8%]
- E. Regular exercise [12%]
- ✔ F. Smoking cessation [67%]
- G. Tight glycemic control [8%]

**Explanation:**



**Explanation:**






One of the objectives for the national Healthy People 2020 initiative is to reduce the prevalence of tobacco smoking in adults from 20% in 2008 to less than 12% by 2020. Tobacco use, especially smoking, is the single most preventable cause of death and disease in the United States. In addition to its numerous direct adverse effects, tobacco smoking also substantially increases the risk for macrovascular (eg, myocardial infarction [MI], stroke) and microvascular (eg, retinopathy, nephropathy) complications of diabetes mellitus. These complications significantly increase mortality, primarily by increasing the risk of MI. The risk of MI-associated mortality begins to decrease immediately upon smoking cessation. However, it takes several tobacco-free years before the risk returns to baseline, especially in patients with a

One of the objectives for the national Healthy People 2020 initiative is to reduce the prevalence of tobacco smoking in adults from 20% in 2008 to less than 12% by 2020. Tobacco use, especially smoking, is the single most preventable cause of death and disease in the United States. In addition to its numerous direct adverse effects, tobacco smoking also substantially increases the risk for macrovascular (eg, myocardial infarction [MI], stroke) and microvascular (eg, retinopathy, nephropathy) complications of diabetes mellitus. These complications significantly increase mortality, primarily by increasing the risk of MI. The risk of MI-associated mortality begins to decrease immediately upon smoking cessation. However, it takes several tobacco-free years before the risk returns to baseline, especially in patients with a long history of smoking. In addition, smokers with diabetes have a higher likelihood of developing hypercholesterolemia, poor glycemic control, and end-stage renal disease. It appears that smoking not only worsens the complications of diabetes but also increases the likelihood of developing diabetes in the first place, thus highlighting the importance of prevention and early cessation.

**(Choice A)** Fecal occult blood testing screens for colorectal cancer and improves mortality through earlier detection. Even so, the mortality benefit is much lower than that achieved by smoking cessation.

**(Choice B)** Prostate-specific antigen is increased in some patients with prostate cancer. However, annual prostate-specific antigen testing has not been shown to improve mortality from prostate cancer.

**(Choice C)** Aspirin use significantly reduces the risk of death from coronary heart disease in diabetic patients, at levels similar to smoking cessation. However, when all-cause mortality is compared, smoking cessation has a greater effect on reducing mortality than daily aspirin therapy.

that achieved by smoking cessation.

**(Choice B)** Prostate-specific antigen is increased in some patients with prostate cancer. However, annual prostate-specific antigen testing has not been shown to improve mortality from prostate cancer.

**(Choice C)** Aspirin use significantly reduces the risk of death from coronary heart disease in diabetic patients, at levels similar to smoking cessation. However, when all-cause mortality is compared, smoking cessation has a greater effect on reducing mortality than daily aspirin therapy.

**(Choices D, E, and G)** Dietary modification and exercise reduce the risk of cardiovascular disease (to a lesser extent than smoking cessation) and help improve glycemic control in diabetic individuals. Although tight glycemic control has been shown to reduce the risk of microvascular complications from diabetes, it does not significantly improve cardiovascular or all-cause mortality.

**Educational objective:**
Smoking cessation is by far the most effective preventive intervention in almost all patients, and this is especially true in those with diabetes.

**References:**

1. How can we best prolong life? Benefits of coronary risk factor reduction in non-diabetic and diabetic subjects.
2. Effect of tight blood glucose control versus conventional control in patients with type 2 diabetes mellitus: a systematic review with meta-analysis of randomized controlled trials.

Time Spent: 11 seconds



Last updated: [9/21/2015]

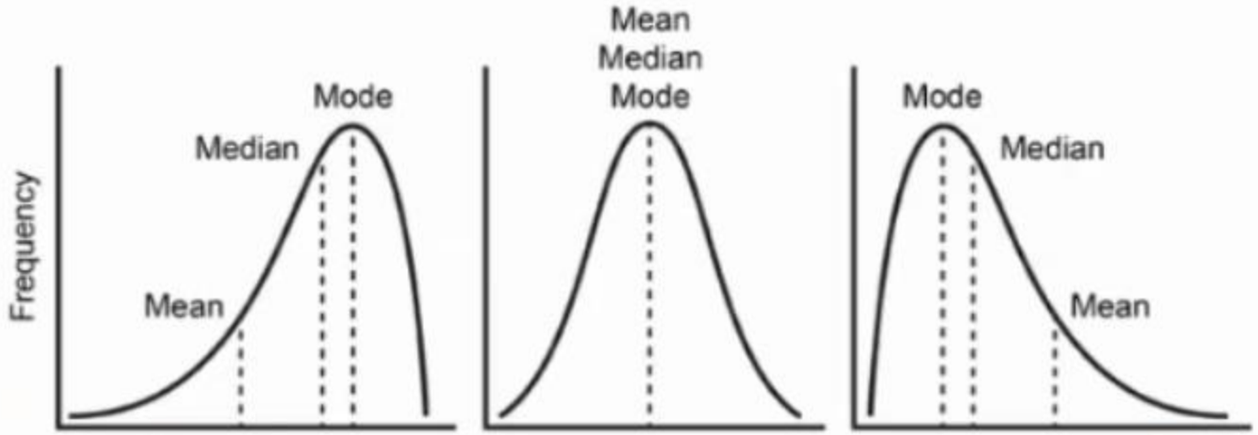
A medical student is conducting a chart review of patients admitted through the emergency department with pancreatitis. As part of the descriptive analysis, blood glucose levels of 800 patients with acute pancreatitis were found to have a strongly positively skewed distribution. Which of the following is most consistent with this finding?

- A. The mean is equal to the median
- B. The mean is equal to the mode
- C. The mean is greater than the median
- D. The median is greater than the mean
- E. The mode is greater than the mean

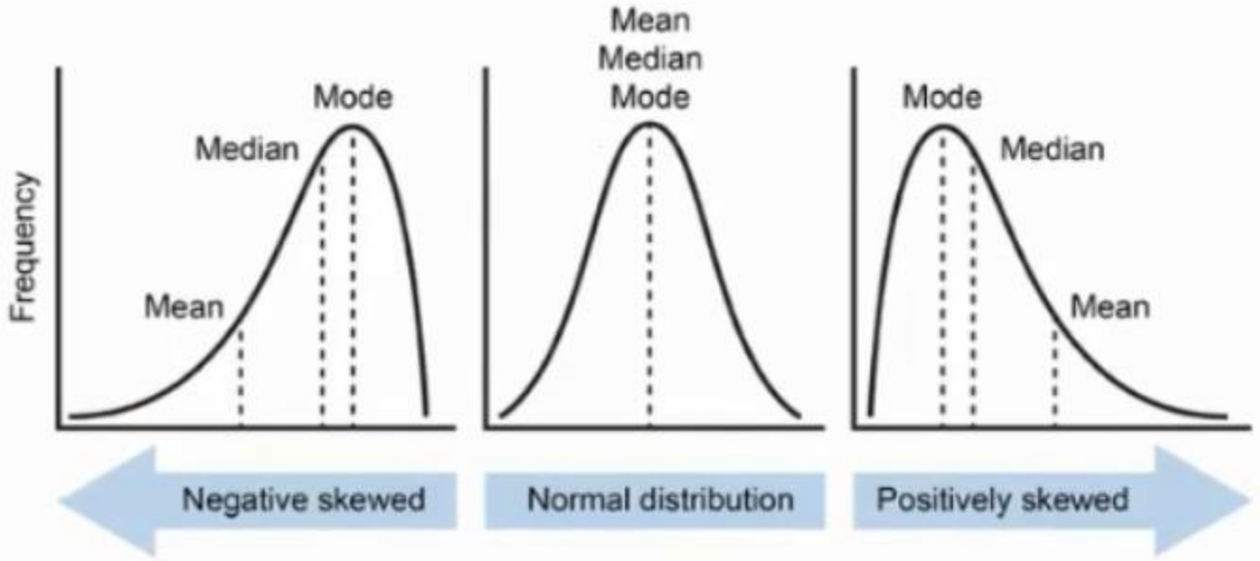
A medical student is conducting a chart review of patients admitted through the emergency department with pancreatitis. As part of the descriptive analysis, blood glucose levels of 800 patients with acute pancreatitis were found to have a strongly positively skewed distribution. Which of the following is most consistent with this finding?

- A. The mean is equal to the median [1%]
- B. The mean is equal to the mode [1%]
- ✔ C. The mean is greater than the median [75%]
- D. The median is greater than the mean [15%]
- E. The mode is greater than the mean [7%]

**Explanation:**



**Explanation:**






Although the symmetrical bell-shaped curve is cited as the prototypical distribution curve, many datasets generated during "real world" statistical analysis have asymmetrical distributions. When a distribution curve is asymmetric, it is either positively or negatively **skewed**. With a positive skew, smaller numbers predominate in the dataset and the long slope of the curve (the "tail") extends in the positive direction (to the right). With a negative skew, larger numbers predominate in the dataset and the long slope of the curve (the "tail") extends in the negative direction (to the left). In general, in a positively skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (mean > median > mode). In such a situation, the median often reflects central tendency better than the mean does.

Negative skewed | Normal distribution | Positively skewed



Although the symmetrical bell-shaped curve is cited as the prototypical distribution curve, many datasets generated during "real world" statistical analysis have asymmetrical distributions. When a distribution curve is asymmetric, it is either positively or negatively **skewed**. With a positive skew, smaller numbers predominate in the dataset and the long slope of the curve (the "tail") extends in the positive direction (to the right). With a negative skew, larger numbers predominate in the dataset and the long slope of the curve (the "tail") extends in the negative direction (to the left). In general, in a positively skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (mean > median > mode). In such a situation, the median often reflects central tendency better than the mean does.

**(Choices A and B)** In normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.

**(Choices D and E)** In general, in a negatively skewed distribution (with the "tail" on the left), the mean is the most shifted in the negative direction, followed by the median and then the mode.

**Educational objective:**
In general, in a positively skewed distribution, the mean is the most shifted in the positive direction (to the right), followed by the median and then the mode.

Time Spent: 12 seconds



Last updated: [1/18/2016]

A new estrogen receptor agonist is being evaluated for the treatment of postmenopausal symptoms. A prospective study shows that the drug increases the risk of deep vein thrombosis (DVT) in treated women who smoke compared to untreated women who smoke, with a relative risk (RR) of 1.70 and p-value of 0.01. In nonsmokers, no increased risk of DVT is evident with use of the drug (RR = 0.96; p-value = 0.68). Which of the following describes this phenomenon?

- A. Confounding
- B. Effect modification
- C. Latent period
- D. Observer bias
- E. Selection bias

A new estrogen receptor agonist is being evaluated for the treatment of postmenopausal symptoms. A prospective study shows that the drug increases the risk of deep vein thrombosis (DVT) in treated women who smoke compared to untreated women who smoke, with a relative risk (RR) of 1.70 and p-value of 0.01. In nonsmokers, no increased risk of DVT is evident with use of the drug (RR = 0.96; p-value = 0.68). Which of the following describes this phenomenon?

- A. Confounding [48%]
- ✔ B. Effect modification [40%]
- C. Latent period [3%]
- D. Observer bias [3%]
- E. Selection bias [7%]

**Explanation:**

**Effect modification** occurs when the effect of an exposure on an outcome is modified by another variable. In this scenario, smoking status modified the effect of the new estrogen receptor agonist (exposure) on deep vein thrombosis (DVT) incidence (outcome). Smokers taking the new estrogen receptor agonist had an increased risk of developing DVT (as the relative risk was >1, indicating higher risk, and the p-value was <0.05, indicating statistical significance), but nonsmokers taking the medication did not (p-value >0.05). Effect modification is not a bias **(Choices D and E)**, as it is not due to flaws in the design or analysis phases of the study. It is a natural phenomenon that should be described, not corrected.

Effect modification is most easily confused with **confounding (Choice A)**, but

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26 • 27 • 28 • 29 • 30 • 31 • 32 • 33 • 34 • 35 • 36 • 37 • 38 • 39 • 40

**Explanation:**

**Effect modification** occurs when the effect of an exposure on an outcome is modified by another variable. In this scenario, smoking status modified the effect of the new estrogen receptor agonist (exposure) on deep vein thrombosis (DVT) incidence (outcome). Smokers taking the new estrogen receptor agonist had an increased risk of developing DVT (as the relative risk was >1, indicating higher risk, and the p-value was <0.05, indicating statistical significance), but nonsmokers taking the medication did not (p-value >0.05). Effect modification is not a bias **(Choices D and E)**, as it is not due to flaws in the design or analysis phases of the study. It is a natural phenomenon that should be described, not corrected.

Effect modification is most easily confused with **confounding (Choice A)**, but stratified analysis (analyzing the cohort as different subgroups) can help distinguish between the two. With effect modification, the different strata will have different measures of association, as seen in this example showing different risks of DVT among smokers compared to nonsmokers. With confounding, stratification usually reveals no significant difference between the strata. For instance, in an analysis of primary school students (of all grade levels), age can be a confounder that muddies the association between shoe size and intelligence. Children with bigger shoe sizes may appear to be more intelligent on initial analysis. However, this association is likely not due to shoe size but rather to age because older children tend to have both bigger feet and more intelligence. When older and younger children are analyzed separately (stratification based on the confounder), the association between shoe size and intelligence disappears.

**(Choice C)** The latent period is the time required for an exposure to begin having an effect. However, there is no information on how latency was handled in this study.

incidence (outcome). Smokers taking the new estrogen receptor agonist had an increased risk of developing DVT (as the relative risk was >1, indicating higher risk, and the p-value was <0.05, indicating statistical significance), but nonsmokers taking the medication did not (p-value >0.05). Effect modification is not a bias **(Choices D and E)**, as it is not due to flaws in the design or analysis phases of the study. It is a natural phenomenon that should be described, not corrected.

Effect modification is most easily confused with **confounding (Choice A)**, but stratified analysis (analyzing the cohort as different subgroups) can help distinguish between the two. With effect modification, the different strata will have different measures of association, as seen in this example showing different risks of DVT among smokers compared to nonsmokers. With confounding, stratification usually reveals no significant difference between the strata. For instance, in an analysis of primary school students (of all grade levels), age can be a confounder that muddies the association between shoe size and intelligence. Children with bigger shoe sizes may appear to be more intelligent on initial analysis. However, this association is likely not due to shoe size but rather to age because older children tend to have both bigger feet and more intelligence. When older and younger children are analyzed separately (stratification based on the confounder), the association between shoe size and intelligence disappears.

**(Choice C)** The latent period is the time required for an exposure to begin having an effect. However, there is no information on how latency was handled in this study.

**Educational objective:**
Effect modification is present when the effect of the main exposure on the outcome is modified by the presence of another variable. Effect modification is not a bias.

Time Spent: 11 seconds



Last updated: [8/9/2015]

In a large population with little migration, the incidence of diabetes mellitus, type II has been stable at 3 cases per 1,000 per year for 30 years. The prevalence of this disease increased progressively over the same period. Which of the following is the most likely explanation of this trend?

- A. High mortality in diabetics
- B. Selective survival
- C. Improved quality of care
- D. Decreased hospitalization rate
- E. Increased diagnostic accuracy

In a large population with little migration, the incidence of diabetes mellitus, type II has been stable at 3 cases per 1,000 per year for 30 years. The prevalence of this disease increased progressively over the same period. Which of the following is the most likely explanation of this trend?

- A. High mortality in diabetics [2%]
- B. Selective survival [4%]
- ✔ C. Improved quality of care [81%]
- D. Decreased hospitalization rate [1%]
- E. Increased diagnostic accuracy [11%]

**Explanation:**

Incidence and prevalence are two important concepts in epidemiology. Incidence is the measure of the appearance of new cases. Prevalence is the measure of those with the disease in the population at a particular point in time. The relationship between these two categories can be demonstrated by the following approximation in a stable population (little migration):

Prevalence = (Incidence) x (Time)

The above vignette describes a disease with a rising prevalence but stable incidence. Such trend can be attributed to factors that prolong the duration of the disease (e.g., improved quality of care).

**(Choice A)** A high mortality rate in diabetics would result in a stable or decreased prevalence.

E. Increased diagnostic accuracy [11%]

**Explanation:**

Incidence and prevalence are two important concepts in epidemiology. Incidence is the measure of the appearance of new cases. Prevalence is the measure of those with the disease in the population at a particular point in time. The relationship between these two categories can be demonstrated by the following approximation in a stable population (little migration):

Prevalence = (Incidence) x (Time)

The above vignette describes a disease with a rising prevalence but stable incidence. Such trend can be attributed to factors that prolong the duration of the disease (e.g., improved quality of care).

**(Choice A)** A high mortality rate in diabetics would result in a stable or decreased prevalence.

**(Choices B and D)** Selective survival and a decreased hospitalization rate are not helpful in explaining the increased prevalence described in this scenario.

**(Choice E)** Increased diagnostic accuracy affects both the prevalence and incidence of a disease.

**Educational Objective:**
An increasing prevalence and stable incidence can be attributed to factors which prolong the duration of a disease (e.g., improved quality of care - this scenario is typical for the USMLE).

Time Spent: 14 seconds



Last updated: [8/19/2015]

A study investigating a new test for diagnosing acute myocardial infarction (AMI) has just been released. The sensitivity of the test is 75% and the specificity is 80%. The study contains 200 patients who actually had an AMI and 400 who actually did not, as determined by the diagnostic gold standard. How many false negatives are in the study?

- A. 50
- B. 80
- C. 120
- D. 150
- E. 320

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

A study investigating a new test for diagnosing acute myocardial infarction (AMI) has just been released. The sensitivity of the test is 75% and the specificity is 80%. The study contains 200 patients who actually had an AMI and 400 who actually did not, as determined by the diagnostic gold standard. How many false negatives are in the study?

- ✔ (●) A. 50 [63%]
- ( ) B. 80 [18%]
- ( ) C. 120 [8%]
- ( ) D. 150 [7%]
- ( ) E. 320 [3%]

**Explanation:**

False negatives describe people who actually have a disease but test negative. In the table below they are represented by square c.

| | Disease positive | Disease negative | |
|---|---|---|---|
| **Test positive** | a<br>True positive | b<br>False positive | a+b |
| **Test negative** | c<br>False negative | d<br>True negative | c+d |

**Explanation:**

False negatives describe people who actually have a disease but test negative. In the table below they are represented by square c.

| | Disease positive | Disease negative | |
|---|---|---|---|
| Test positive | a<br>True positive | b<br>False positive | a+b |
| Test negative | c<br>False negative | d<br>True negative | c+d |
| | a+c | b+d | |

Sensitivity = a/(a+c)
Specificity = d/(b+d)



The number of false negatives can be calculated using the test's sensitivity and total number of patients with AMI. The sensitivity of a test refers to the number of true positives divided by the total number of patients with the disease (a / [a + c]). In this case, the sensitivity is 75% and the total number of patients with AMI (a + c) equals 200. Using this information, the number of true positives (a) can be calculated:

Sensitivity = a / (a + c)
0.75 = a / 200
a = 150

Since the total number of patients with the disease (a + c) equals 200 and the number of true positives (a) equals 150, the number of false negatives (c) is 50.

For the sake of thoroughness, the two-by-two table can be completed as follows. Specificity refers to the number of true negatives divided by all people without the disease (d / [b + d]). Repeating the same exercise as above gives the completed table:

Specificity = d / (b + d)
0.8 = d / 400
d = 320
b = 400 - 320 = 80

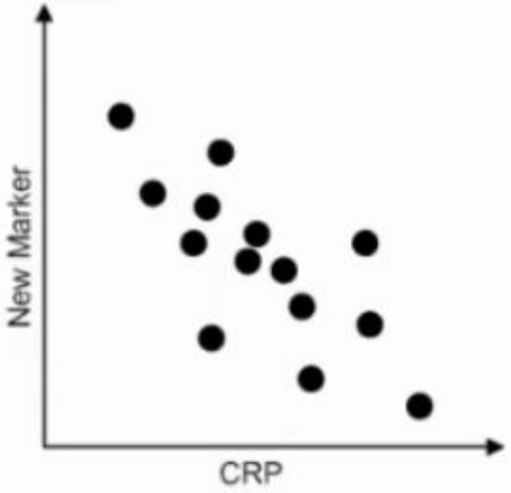
| | AMI | No AMI | |
|---|---|---|---|
| Test positive | 150 | 80 | 230 |
| Test negative | 50 | 320 | 370 |
| | 200 | 400 | |

The number of true positives and false negatives can be more quickly calculated using the following equations:

True positives = (Sensitivity) * (Number of patients actually with the disease)
False negatives = (1 - Sensitivity) * (Number of patients actually with the disease)

**Educational objective:**
Sensitivity is the number of true positives divided by the total number of subjects actually with the disease.

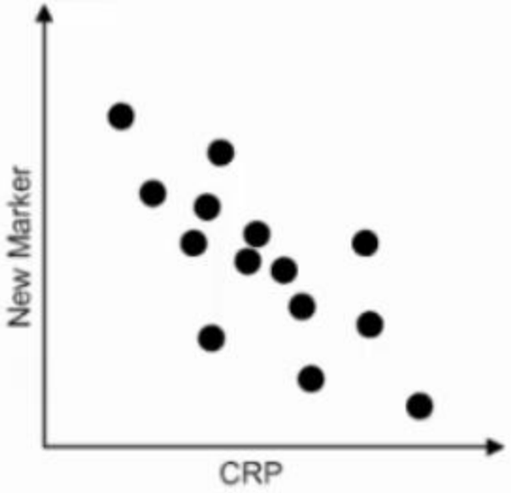
Inflammatory biological markers are clinically useful in a number of different ways, one being the assessment of disease activity in conditions such as systemic lupus erythematosus (SLE). A new inflammatory marker is being investigated in patients with active SLE flares. When the blood level of the new marker (in mg/L) is plotted against the C-reactive protein (CRP) level (also in mg/L), the following plot is obtained:




Based on the plot, the correlation coefficient between the 2 variables is closest to which of the following values?

- A. +0.8
- B. +0.2
- C. 0
- D. −0.2
- E. −0.8

Inflammatory biological markers are clinically useful in a number of different ways, one being the assessment of disease activity in conditions such as systemic lupus erythematosus (SLE). A new inflammatory marker is being investigated in patients with active SLE flares. When the blood level of the new marker (in mg/L) is plotted against the C-reactive protein (CRP) level (also in mg/L), the following plot is obtained:

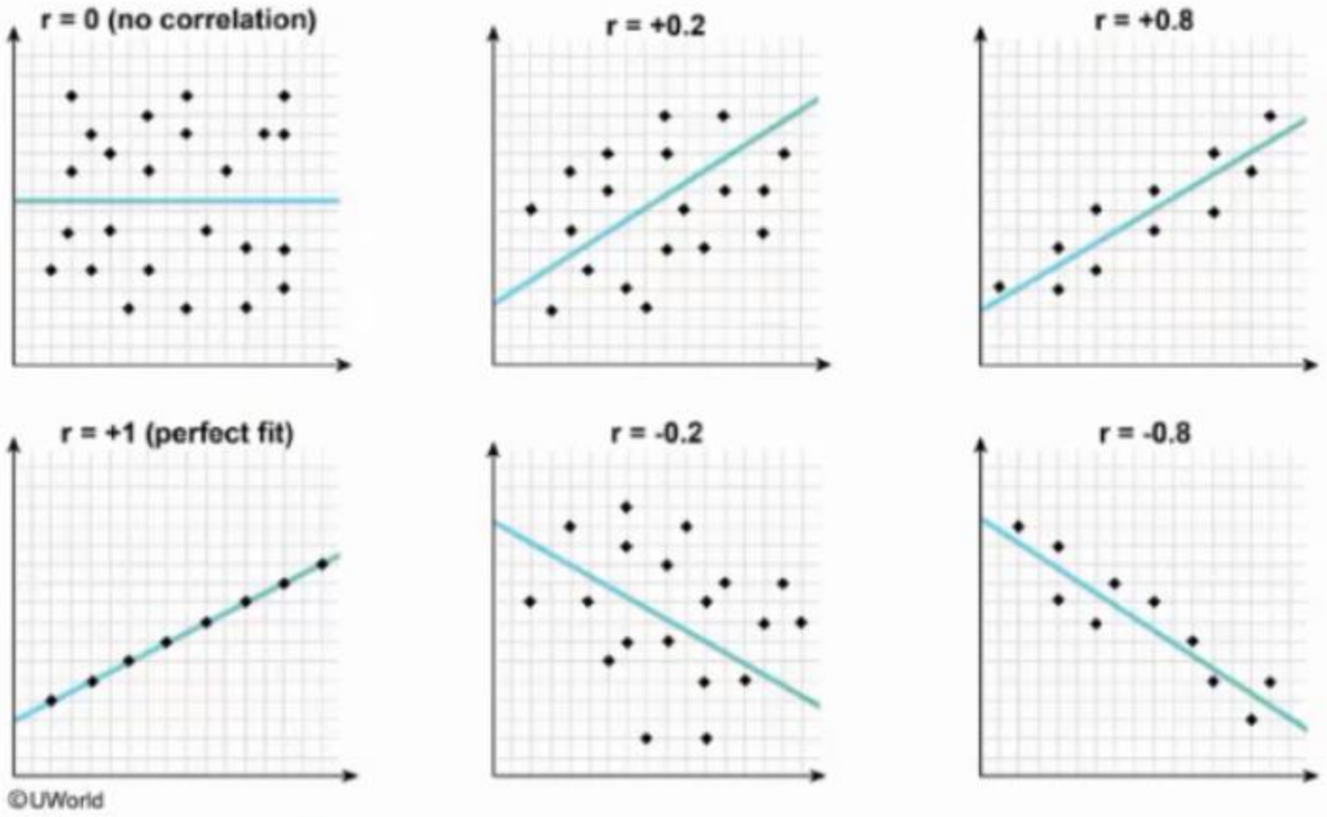



Based on the plot, the correlation coefficient between the 2 variables is closest to which of the following values?

- A. +0.8 [6%]
- B. +0.2 [8%]
- C. 0 [8%]
- D. −0.2 [13%]
- ✔ E. −0.8 [64%]

**Explanation:**

## Correlation coefficients




Scatter plots are useful for crude data analysis. If a **linear** association is present between 2 variables, a **correlation coefficient** (r) mathematically describes how well a "line of best fit" (blue line in Figure) would correspond to the data points plotted. The value of r ranges from **−1 to +1** and describes 2 important characteristics of an association: the strength and the polarity. The closer the r value

Scatter plots are useful for crude data analysis. If a **linear** association is present between 2 variables, a **correlation coefficient** (r) mathematically describes how well a "line of best fit" (blue line in Figure) would correspond to the data points plotted. The value of r ranges from **−1 to +1** and describes 2 important characteristics of an association: the strength and the polarity. The closer the r value is to its margins [−1, 1], the stronger the association.

An increase in C-reactive protein (CRP) level is associated with a decrease in the new marker's level; therefore, r <0. Although the scatter plot does not demonstrate a perfect linear arrangement, it does show a reasonably strong linear association as there is minimal variation along the line of best fit (points are relatively close to the line). Therefore, the most appropriate answer among the options given is r = −0.8.

**(Choices A and B)** In this example, r cannot be >0 as an increase in CRP level is associated with a decrease in the new marker's level.

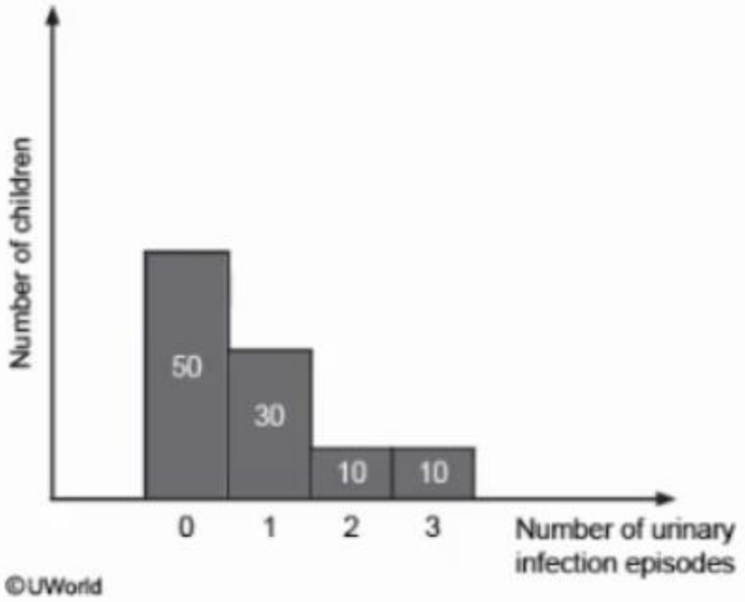
**(Choice C)** An r = 0 indicates that there is no association (ie, a random distribution).

**(Choice D)** An r = −0.2 would also indicate a negative association; however, there would be more variation along the line of best fit (as seen in the Figure).

The value of r is **not** the slope of the line of best fit. For instance, assuming a data set with a maximum positive association (ie, all data points forming a straight line with a positive slope), the line of best fit will always have r = +1, even if its actual slope is 0.2, 5.5, or any other positive value (as seen in the Figure).

**Educational objective:**
The correlation coefficient (r) ranges from −1 to +1 and describes the strength and polarity of a linear association.

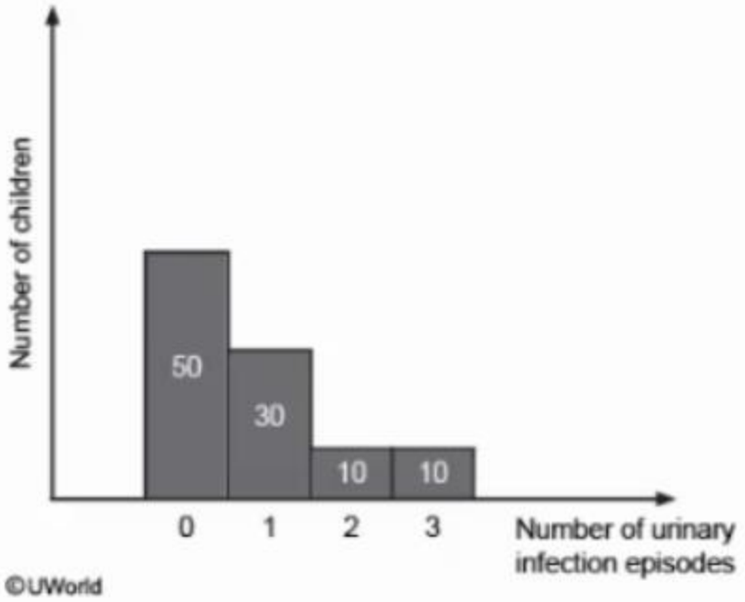
A sample of children age 2-5 years is chosen at random from an outpatient clinic. The number of urinary tract infections (UTIs) over a 1-year period is given in the figure below.




What was the average number of UTI episodes over 1 year for a child in this sample?

- A. Between 0 and 1
- B. 1
- C. Between 1 and 2
- D. 2
- E. Between 2 and 3

A sample of children age 2-5 years is chosen at random from an outpatient clinic. The number of urinary tract infections (UTIs) over a 1-year period is given in the figure below.




What was the average number of UTI episodes over 1 year for a child in this sample?

- ✔ A. Between 0 and 1 [74%]
- B. 1 [4%]
- C. Between 1 and 2 [17%]
- D. 2 [2%]
- E. Between 2 and 3 [3%]

- B. 1 [4%]
- C. Between 1 and 2 [17%]
- D. 2 [2%]
- E. Between 2 and 3 [3%]

**Explanation:**

This bar graph illustrates the frequency of urinary tract infection (UTI) episodes in the sample. According to the figure, 50 children had 0 UTIs, 30 children had 1 UTI, 10 had 2 UTIs, and 10 had 3 UTIs.

In general, the average (or mean) of a dataset is the sum of the values divided by the total number of values. In this example, the average number of UTI episodes per child is the sum (ie, total number) of UTIs divided by the total sample size (ie, total number of children). The total sample size is: 50 + 30 + 10 + 10 = 100 children. The total number of UTIs is: (0 × 50) + (1 × 30) + (2 × 10) + (3 × 10) = 0 + 30 + 20 + 30 = 80 UTIs. The average is obtained by dividing the total number of UTIs (80) by the total sample size (100). Therefore, the average number of UTI episodes per year = 80/100 = 0.8 UTIs for a child in this sample.

In conclusion, children age 2-5 years in this clinic experienced between 0 and 1 UTI over 1 year on average.

**Educational objective:**
The average (or mean) of a dataset of values is the sum of the values divided by the total number of values.

Time Spent: 13 seconds



Last updated: [1/19/2016]

A 65-year-old man with a history of congestive heart failure is hospitalized with chest pain and hypotension requiring admission to the cardiac care unit. An intra-arterial line is placed for direct blood pressure monitoring. Consecutive readings of his intra-arterial blood pressure are 75, 110, 80, 90, 75, and 110 mm Hg. Which of the following represents the median of these blood pressure readings?

- A. 80 mm Hg
- B. 85 mm Hg
- C. 90 mm Hg
- D. 100 mm Hg
- E. 110 mm Hg

A 65-year-old man with a history of congestive heart failure is hospitalized with chest pain and hypotension requiring admission to the cardiac care unit. An intra-arterial line is placed for direct blood pressure monitoring. Consecutive readings of his intra-arterial blood pressure are 75, 110, 80, 90, 75, and 110 mm Hg. Which of the following represents the median of these blood pressure readings?

- A. 80 mm Hg [3%]
- ✔ B. 85 mm Hg [88%]
- C. 90 mm Hg [8%]
- D. 100 mm Hg [0%]
- E. 110 mm Hg [1%]

**Explanation:**

It is important to understand the difference between the median, mean, and mode, which are 3 measures of central tendency. The **median** of an ordered dataset is the number that separates the right half of the data from the left half. The dataset MUST be ordered before the median is determined. If the number of observations in the dataset is even, finding the median requires adding the middle 2 values together and dividing that sum by 2.

The ordered dataset in this case is {75, 75, **80, 90**, 110, 110}. There are 6 observations, which is an even number of observations. The median value splits the dataset in half; it lies between 80 and 90 (3 values on the left and 3 values on the right). The median is (80 + 90) / 2 = 85 mm Hg.

**Explanation:**

It is important to understand the difference between the median, mean, and mode, which are 3 measures of central tendency. The **median** of an ordered dataset is the number that separates the right half of the data from the left half. The dataset MUST be ordered before the median is determined. If the number of observations in the dataset is even, finding the median requires adding the middle 2 values together and dividing that sum by 2.

The ordered dataset in this case is {75, 75, **80, 90**, 110, 110}. There are 6 observations, which is an even number of observations. The median value splits the dataset in half; it lies between 80 and 90 (3 values on the left and 3 values on the right). The median is (80 + 90) / 2 = 85 mm Hg.

Now, assume one of the values is missing and the ordered dataset includes the following 5 observations (odd number of observations): {75, 80, **90**, 110, 110}. In this case, the median value would be 90 mm Hg, which splits the dataset in half (2 values on the left and 2 values on the right).

**(Choices A and D)** The values 80 and 100 mm Hg are not measures of the center of the dataset.

**(Choice C)** To find the **mean** of a dataset, add all of the observations and divide that sum by the number of observations. In this case, the mean is (75 + 75 + 80 + 90 + 110 + 110) / 6 = 90 mm Hg.

**(Choice E)** Another measure of the center of a dataset is the mode. Finding the **mode** is easy because it is the most frequent value of a dataset. In our scenario, the dataset is "bimodal" because 75 and 110 mm Hg are each listed twice.

number that separates the right half of the data from the left half. The dataset must be ordered before the median is determined. If the number of observations in the dataset is even, finding the median requires adding the middle 2 values together and dividing that sum by 2.

The ordered dataset in this case is {75, 75, **80, 90**, 110, 110}. There are 6 observations, which is an even number of observations. The median value splits the dataset in half; it lies between 80 and 90 (3 values on the left and 3 values on the right). The median is (80 + 90) / 2 = 85 mm Hg.

Now, assume one of the values is missing and the ordered dataset includes the following 5 observations (odd number of observations): {75, 80, **90**, 110, 110}. In this case, the median value would be 90 mm Hg, which splits the dataset in half (2 values on the left and 2 values on the right).

**(Choices A and D)** The values 80 and 100 mm Hg are not measures of the center of the dataset.

**(Choice C)** To find the **mean** of a dataset, add all of the observations and divide that sum by the number of observations. In this case, the mean is (75 + 75 + 80 + 90 + 110 + 110) / 6 = 90 mm Hg.

**(Choice E)** Another measure of the center of a dataset is the mode. Finding the **mode** is easy because it is the most frequent value of a dataset. In our scenario, the dataset is "bimodal" because 75 and 110 mm Hg are each listed twice.

**Educational objective:**
The median is the value that is located in the precise center of an ordered dataset. It separates the right half of the data from the left half.

Time Spent: 14 seconds



Last updated: [1/19/2016]

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 and 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?

- A. The number of incident cases will decrease, the number of prevalent cases will decrease
- B. The number of incident cases will increase, the number of prevalent cases will not change
- C. The number of incident cases will decrease, the number of prevalent cases will increase
- D. The number of incident cases will not change, the number of prevalent cases will increase
- E. The number of incident cases will not change, the number of prevalent cases will not change

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 and 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?

- A. The number of incident cases will decrease, the number of prevalent cases will decrease [1%]
- B. The number of incident cases will increase, the number of prevalent cases will not change [2%]
- C. The number of incident cases will decrease, the number of prevalent cases will increase [2%]
- ✔ D. The number of incident cases will not change, the number of prevalent cases will increase [91%]
- E. The number of incident cases will not change, the number of prevalent cases will not change [4%]

**Explanation:**

**Incident cases** represent **new cases** diagnosed in a given period of time.
**Prevalent cases** are the **total number of cases** (both old and new) at a particular point in time. For a non-transmissible condition such as chronic lymphocytic

**Explanation:**

**Incident cases** represent **new cases** diagnosed in a given period of time. **Prevalent cases** are the **total number of cases** (both old and new) at a particular point in time. For a non-transmissible condition such as chronic lymphocytic leukemia (CLL), the number of incident cases will **not** be changed by any kind of treatment because the disease has already developed when the treatment is started (and treatment of existing patients will not prevent new patients from developing CLL). Any treatment that prolongs survival but does not cure the disease will **increase** the number of prevalent cases due to an increase in the number of afflicted (but still living) individuals over time.

Differences in terminology may lead to confusion. Incident cases and prevalent cases are often referred to as simply "incidence" and "prevalence." Incidence can also represent the number of new cases divided by the population at risk over a period of time and can be converted into a rate (eg, annual incidence per 10,000 population). Prevalence typically refers to point prevalence (total number of cases at a particular point in time) as opposed to period prevalence (over a period of time). The specific meaning is often clear from the context.

**Educational objective:**
Incident cases represent new cases diagnosed in a given period of time. Prevalent cases are the total number of cases (both old and new) at a particular point in time. Any treatment that prolongs survival but does not cure the disease will increase prevalence due to an increase in the number of afflicted (but still living) individuals over time.

Time Spent: 12 seconds



Last updated: [11/4/2015]

Physician researchers are evaluating the efficacy of a new lipid-lowering drug being marketed as Superstatin. The drug manufacturer claims that the drug is more effective than existing hypolipidemic agents in preventing myocardial infarctions. Results of a 5-year randomized control trial are shown below.

| | Number of patients treated with Superstatin | Number of patients treated with control medication |
|---|---|---|
| Myocardial infarction | 10 | 25 |
| No myocardial infarction | 990 | 975 |

Compared to the control medication, how many patients need to be treated with Superstatin to prevent one additional myocardial infarction?

- A. 1
- B. 2
- C. 2.5
- D. 15
- E. 40
- F. 67
- G. 100

Physician researchers are evaluating the efficacy of a new lipid-lowering drug being marketed as Superstatin. The drug manufacturer claims that the drug is more effective than existing hypolipidemic agents in preventing myocardial infarctions. Results of a 5-year randomized control trial are shown below.

| | Number of patients treated with Superstatin | Number of patients treated with control medication |
|---|---|---|
| Myocardial infarction | 10 | 25 |
| No myocardial infarction | 990 | 975 |

Compared to the control medication, how many patients need to be treated with Superstatin to prevent one additional myocardial infarction?

- A. 1 [5%]
- B. 2 [4%]
- C. 2.5 [17%]
- D. 15 [13%]
- E. 40 [7%]
- ✔ F. 67 [44%]
- G. 100 [10%]

**Explanation:**

## Explanation:

| Common measures of therapeutic efficacy | | |
|---|---|---|
| **Term** | **Definition** | **Calculation** |
| **Absolute Risk Reduction (ARR)** | Percentage indicating the actual difference in event rate between control and treatment groups. | ARR = Control Rate – Treatment Rate |
| **Relative Risk Reduction (RRR)** | Percentage indicating relative reduction in the treatment event rate compared to the control group. | RRR = ARR/Control Rate |
| **Relative Risk (RR)** | Ratio of the probability of an event occurring in the treatment group compared to the control group. | RR = Treatment Rate/Control Rate |
| **Number Needed to Treat (NNT)** | Number of individuals that need to be treated to prevent a negative outcome in one patient. | NNT = 1/ARR |



The number needed to treat (NNT) represents the number of patients that need to

| | | |
|---|---|---|
| **Number Needed to Treat (NNT)** | Number of individuals that need to be treated to prevent a negative outcome in one patient. | NNT = 1/ARR |



The number needed to treat (NNT) represents the number of patients that need to be treated with a medication in order to prevent an additional negative outcome. NNT is calculated by dividing 1 by the absolute risk reduction (ARR). In this example, ARR is calculated by subtracting the event rate in the treatment group (10/1000= 0.01 or 1%) from the event rate in the control group (25/1000 = 0.025 or 2.5%). Performing the calculation (0.025 - 0.01) results in an ARR value of 0.015 or 1.5%. Finally, dividing 1 by 0.015 gives an NNT of 66.6. Rounding up the result to the nearest whole number, 67 patients need to be treated with Superstatin to prevent an additional myocardial infarction.

A similar concept to NNT is the number needed to harm (NNH), which usually applies to medication side effects. In both cases, lower values of NNT or NNH imply that an intervention is more beneficial (NNT) or harmful (NNH). For instance, if a third hypolipidemic medication had a NNT of 40, it would be more beneficial than Superstatin as only 40 patients would need to be treated to prevent an additional myocardial infarction.

**Educational objective:**
The number needed to treat (NNT) is the number of patients that need to be treated with a medication to avoid a negative outcome in one patient. NNT is calculated by dividing 1 by the absolute risk reduction. Lower NNT values represent more beneficial treatments.

Time Spent: 12 seconds



Last updated: [12/19/2015]

Officials of a large community hospital report an increased incidence of acute myelogenous leukemia (AML) among children age 5-12 years. They observe that some households in the community are exposed to chemical waste from a nearby factory and worry that exposure to this waste is responsible for the increased incidence of AML. A case-control study is designed to evaluate the hospital officials' claim that exposure to chemical waste increases the risk for developing AML in childhood. Which of the following populations is most likely to function as the control group?

- A. Children who do not have AML and are exposed to chemical waste
- B. Children who do not have AML and are not exposed to chemical waste
- C. Children who do not have AML, regardless of exposure status to chemical waste
- D. Children who have AML and are exposed to chemical waste
- E. Children who have AML and are not exposed to chemical waste
- F. Children who have AML, regardless of exposure status to chemical waste

Officials of a large community hospital report an increased incidence of acute myelogenous leukemia (AML) among children age 5-12 years. They observe that some households in the community are exposed to chemical waste from a nearby factory and worry that exposure to this waste is responsible for the increased incidence of AML. A case-control study is designed to evaluate the hospital officials' claim that exposure to chemical waste increases the risk for developing AML in childhood. Which of the following populations is most likely to function as the control group?

- A. Children who do not have AML and are exposed to chemical waste [20%]
- B. Children who do not have AML and are not exposed to chemical waste [23%]
- ✔ C. Children who do not have AML, regardless of exposure status to chemical waste [28%]
- D. Children who have AML and are exposed to chemical waste [2%]
- E. Children who have AML and are not exposed to chemical waste [24%]
- F. Children who have AML, regardless of exposure status to chemical waste [3%]

**Explanation:**

A case-control study is the most appropriate study design for evaluating the hospital officials' claim. This is because the disease is known (acute myelogenous leukemia or AML), but a retrospective risk factor (chemical waste exposure) needs to be evaluated. In case-control studies, 2 groups of subjects are created: cases

difference in exposure between the 2 groups (eg, if the exposure to chemical waste is significantly higher among patients with AML compared to patients without AML), then it is likely that the retrospective risk factor in question is indeed a true risk factor for disease development.

In this example, children who do not have AML should be used as controls, given that AML is the outcome of interest. They should be selected regardless of exposure to the chemical waste. Selecting control subjects based on exposure status would be inappropriate because comparing the frequency of exposure between the case and control groups is an important part of case-control study analysis.

**(Choices A and B)** Optimal selection of control subjects provides an unbiased estimation of exposure frequency amongst the non-diseased population. Skewing the control group (eg, through non-random selection) so that it contains a higher or lower exposure frequency could lead to erroneous results. To decrease the effects of confounding, independent variables not being tested (eg, age and sex) are often specifically selected to be the same (matched) between case and control groups. However, matching must be carefully performed so as not to introduce selection bias.

**(Choices D, E, and F)** AML is the outcome of interest; therefore, children who have AML can only be used to form the cases group and cannot be used as controls.

**Educational objective:**
Selection of control subjects in case-control studies is intended to provide an accurate estimation of exposure frequency among the non-diseased general population. Cases and controls are often matched to decrease confounding. However, matching must be carefully performed so as to not introduce selection bias.

Time Spent: 16 seconds



Last updated: [1/19/2016]

A study determines that the mean blood cholesterol level is 195 mg/dL in 200 non-diabetic hospitalized patients and 210 mg/dL in 180 diabetic hospitalized patients. The probability that the observed difference is due to chance alone is reported to be 5%. There is also a 20% probability of concluding that there is no difference in blood cholesterol level when there is one in reality. What is the power of the study?

- A. 0.05
- B. 0.20
- C. 0.50
- D. 0.80
- E. 0.95

A study determines that the mean blood cholesterol level is 195 mg/dL in 200 non-diabetic hospitalized patients and 210 mg/dL in 180 diabetic hospitalized patients. The probability that the observed difference is due to chance alone is reported to be 5%. There is also a 20% probability of concluding that there is no difference in blood cholesterol level when there is one in reality. What is the power of the study?

- A. 0.05 [8%]
- B. 0.20 [7%]
- C. 0.50 [4%]
- ✔ D. 0.80 [71%]
- E. 0.95 [11%]

**Explanation:**

The **power** of a study is the ability of a study to detect a difference between groups when such a difference truly exists. Power is related to **type II error (β)**, which is the probability of concluding there is no difference between groups when one truly exists. Mathematically, power is given by:

**Power = 1 – β**

In this example, the power of the study is the probability of detecting a difference in blood cholesterol level between diabetics and non-diabetics if there is a real difference. The probability of concluding that there is no difference in blood cholesterol level when in reality there is one is given as 20%; this corresponds to the definition of β (ie, β = 0.20 in this example). Therefore:

**Explanation:**

The **power** of a study is the ability of a study to detect a difference between groups when such a difference truly exists. Power is related to **type II error (β)**, which is the probability of concluding there is no difference between groups when one truly exists. Mathematically, power is given by:

**Power = 1 – β**

In this example, the power of the study is the probability of detecting a difference in blood cholesterol level between diabetics and non-diabetics if there is a real difference. The probability of concluding that there is no difference in blood cholesterol level when in reality there is one is given as 20%; this corresponds to the definition of β (ie, β = 0.20 in this example). Therefore:

Power = 1 – β = 1 – 0.20 = 0.80

**(Choice A)** Type I error (α) describes the probability of seeing a difference when there is no difference in reality. The value of α is generally compared to the probability that the observed difference is due to chance alone (a simplified explanation of the p-value). In this example, the probability that the observed difference between diabetic and nondiabetic patients is due to chance alone is given as 5% (0.05).

**(Choices B and E)** Type II error (β) is 0.20, as explained above. The value 0.95 corresponds to (1 – α), but power is given by (1 – β).

**Educational objective:**
The power of a study indicates the probability of seeing a difference when there is one. The formula is Power = 1 – β, where β is the type II error rate.

A new test to diagnose urinary tract infections in women is being evaluated. The comparison gold standard is positive urine dipstick plus urine culture. Results of the study are given below.

| | Positive urine dipstick & culture | Negative urine dipstick & culture |
|---|---|---|
| Test positive | 60 | 20 |
| Test negative | 140 | 180 |

What is the new test's specificity?

- A. 10%
- B. 30%
- C. 50%
- D. 70%
- E. 90%

A new test to diagnose urinary tract infections in women is being evaluated. The comparison gold standard is positive urine dipstick plus urine culture. Results of the study are given below.

| | Positive urine dipstick & culture | Negative urine dipstick & culture |
|---|---|---|
| Test positive | 60 | 20 |
| Test negative | 140 | 180 |

What is the new test's specificity?

- A. 10% [5%]
- B. 30% [8%]
- C. 50% [3%]
- D. 70% [4%]
- ✔ E. 90% [79%]

**Explanation:**

| | Positive condition | Negative condition |
|---|---|---|

**Explanation:**

| | Positive condition | Negative condition | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |
| | Sensitivity = TP/(TP+FN) | Specificity = TN/(TN+FP) | |

**FN** = false negative; **FP** = false positive; **TN** = true negative; **TP** = true positive; **NPV** = negative predictive value; **PPV** = positive predictive value.



The **specificity** of a test refers to its ability to correctly identify individuals without the disease; it is the number of true negatives (TN) divided by all those without the disease. Specificity should be high in **confirmatory** tests to decrease the number of false positives (FP). Using a generic 2×2 (contingency) table, specificity can be calculated from: TN / (TN + FP). In this case:

Specificity = TN / (TN + FP) = 180 / (180 + 20) = 180/200 = 0.90 (or 90%)

**(Choice A)** In this example, FP / (TN + FP) = 20 / (180 + 20) = 20/200 = 10%. This value is called the false positive rate, which is equal to (1 – specificity). It represents the probability of a false positive test result in a patient who in reality is free of disease.

**(Choices B and D)** Sensitivity (a test's ability to correctly identify individuals with the

The **specificity** of a test refers to its ability to correctly identify individuals without the disease; it is the number of true negatives (TN) divided by all those without the disease. Specificity should be high in **confirmatory** tests to decrease the number of false positives (FP). Using a generic 2×2 (contingency) table, specificity can be calculated from: TN / (TN + FP). In this case:

Specificity = TN / (TN + FP) = 180 / (180 + 20) = 180/200 = 0.90 (or 90%)

**(Choice A)** In this example, FP / (TN + FP) = 20 / (180 + 20) = 20/200 = 10%. This value is called the false positive rate, which is equal to (1 – specificity). It represents the probability of a false positive test result in a patient who in reality is free of disease.

**(Choices B and D)** Sensitivity (a test's ability to correctly identify individuals with the disease) can be given by TP / (TP + FN), where TP is true positives and FN is false negatives. In this example, sensitivity = TP / (TP + FN) = 60 / (60 + 140) = 60/200 = 0.30 (or 30%). The false negative rate (= 1 – sensitivity) is given by FN / (TP + FN) = 140 / (60 + 140) = 140/200 = 0.70 (or 70%). It is the probability of a false negative test result in a patient who in reality has the disease.

**(Choice C)** From a total of 400 patients (= 60 + 140 + 20 + 180), there are 200 patients (= 60 + 140) who have a urinary tract infection (UTI) based on the gold standard. The prevalence of UTIs is therefore 200/400 = 0.50 (or 50%).

**Educational objective:**
The specificity of a test is its ability to correctly identify individuals without the disease. Specificity should be high in confirmatory tests to decrease false positives.

Time Spent: 9 seconds



Last updated: [8/30/2015]

A 34-year-old man is admitted to the hospital with acute chest pain. An electrocardiogram is obtained in the emergency department and shows ST segment elevation in leads II, III and avF. A sample of blood is taken from the patient. A new test is used to measure the plasma homocysteine level 3 times with this blood sample. The results are 11.2 µmol/L, 13.5 µmol/L, and 17.1 µmol/L (laboratory reference range: 4-14 µmol/L). These results suggest that concerns should be raised about which of the following with the new test?

- A. Accuracy
- B. Reliability
- C. Sensitivity
- D. Specificity
- E. Validity

A 34-year-old man is admitted to the hospital with acute chest pain. An electrocardiogram is obtained in the emergency department and shows ST segment elevation in leads II, III and avF. A sample of blood is taken from the patient. A new test is used to measure the plasma homocysteine level 3 times with this blood sample. The results are 11.2 µmol/L, 13.5 µmol/L, and 17.1 µmol/L (laboratory reference range: 4-14 µmol/L). These results suggest that concerns should be raised about which of the following with the new test?

- A. Accuracy [21%]
- ✔ B. Reliability [70%]
- C. Sensitivity [2%]
- D. Specificity [2%]
- E. Validity [5%]

**Explanation:**

**Reliability & validity**

Unreliable & invalid

Unreliable & valid

Reliable & invalid

Reliable & valid

A **reliable** test is **reproducible**; it gives similar or very close results on repeat measurements (bottom half of figure). In this example, repeat measurements of the same sample yielded markedly different results; therefore, the new test is not very reliable. As an example, in a clinical laboratory, reliability is typically quantified in terms of the coefficient of variation (standard deviation divided by mean of repeated measurements), generally expressed as a percentage. The coefficient of variation, which is not the same as the laboratory reference range, can be used to compare tests with different normal values and/or units. The coefficient of variation should be calculated for this test and compared to that of other tests measuring plasma homocysteine.

**(Choices A and E) Validity (accuracy)** is defined as a test's ability to measure what it is supposed to measure. To evaluate the validity of a new test, its results should be compared to those obtained with the "gold standard" test on the same individual. In this case, as there were no test results obtained using the "gold standard," the validity/accuracy of the test cannot be determined. A test can be highly reliable (ie, gives very similar results on repeat measurements) but invalid (ie, the measurements are all incorrect compared to the gold standard).

**(Choices C and D)** Sensitivity is defined as a test's ability to identify the true presence of disease, whereas specificity is defined as a test's ability to identify the true absence of disease. The reliability of a test is unrelated to its sensitivity or specificity.

**Educational objective:**
A reliable test is reproducible in that it gives similar results on repeat measurements. Reliability is maximal when random error is minimal.

Time Spent: 9 seconds



Last updated: [8/9/2015]

A study is conducted to assess the effect of beta-blocker therapy in patients with acute myocardial infarction (MI). The study found that 20 patients out of 90 who took a beta-blocker during the week prior to an MI developed a major arrhythmia during hospitalization. The study also determined that 30 patients out of 70 who did not take any beta-blocker developed a major arrhythmia. What was the odds ratio of developing major arrhythmia in patients who took beta-blockers compared to those who did not take beta-blockers?

- A. (20 × 40) / (70 × 30)
- B. (20 × 70) / (30 × 40)
- C. (20 × 30) / (70 × 40)
- D. (20/50) / (70/110)
- E. (20/90) / (30/70)

A study is conducted to assess the effect of beta-blocker therapy in patients with acute myocardial infarction (MI). The study found that 20 patients out of 90 who took a beta-blocker during the week prior to an MI developed a major arrhythmia during hospitalization. The study also determined that 30 patients out of 70 who did not take any beta-blocker developed a major arrhythmia. What was the odds ratio of developing major arrhythmia in patients who took beta-blockers compared to those who did not take beta-blockers?

- ✔ (●) A. (20 × 40) / (70 × 30) [50%]
- ( ) B. (20 × 70) / (30 × 40) [6%]
- ( ) C. (20 × 30) / (70 × 40) [7%]
- ( ) D. (20/50) / (70/110) [2%]
- ( ) E. (20/90) / (30/70) [35%]

**Explanation:**

The odds of an event is defined as the probability of that event divided by 1 minus the probability of the event: Pr[Event] / (1 − Pr[Event]). The **odds ratio (OR)** is a measure of association between an exposure and an outcome. In this case, it represents the odds that an outcome (eg, major arrhythmia) occurred in the presence of a particular exposure (eg, beta-blocker) compared to the odds that the outcome occurred in the absence of that exposure. Using a standard contingency table (with exposures in the rows and outcomes in the columns), the mathematical shortcut for OR is given by:

OR = ad/bc

**Explanation:**

The odds of an event is defined as the probability of that event divided by 1 minus the probability of the event: Pr[Event] / (1 − Pr[Event]). The **odds ratio (OR)** is a measure of association between an exposure and an outcome. In this case, it represents the odds that an outcome (eg, major arrhythmia) occurred in the presence of a particular exposure (eg, beta-blocker) compared to the odds that the outcome occurred in the absence of that exposure. Using a standard contingency table (with exposures in the rows and outcomes in the columns), the mathematical shortcut for OR is given by:

OR = ad/bc

For this example, the contingency table can be completed as follows:

| | Major arrhythmia | No arrhythmia | |
|---|---|---|---|
| Beta-blocker | a = 20 | b = 70 | 90 |
| No beta-blocker | c = 30 | d = 40 | 70 |
| | 50 | 110 | 160 |

Here, a + b = 90 (total number of patients on beta-blocker) and a = 20 (number of patients on beta-blocker who developed arrhythmia), so b = 90 – 20 = 70. Similarly, c + d = 70 (total number of patients not taking beta-blocker) and c = 30 (number of patients not taking beta-blocker who developed arrhythmias), so d = 70 – 30 = 40.

Therefore, the OR = ad/bc = (20 × 40) / (70 × 30) = 0.38. This means that the odds

| | arrhythmia | arrhythmia | |
|---|---|---|---|
| Beta-blocker | a = 20 | b = 70 | 90 |
| No beta-blocker | c = 30 | d = 40 | 70 |
| | 50 | 110 | 160 |

Here, a + b = 90 (total number of patients on beta-blocker) and a = 20 (number of patients on beta-blocker who developed arrhythmia), so b = 90 – 20 = 70. Similarly, c + d = 70 (total number of patients not taking beta-blocker) and c = 30 (number of patients not taking beta-blocker who developed arrhythmias), so d = 70 – 30 = 40.

Therefore, the OR = ad/bc = (20 × 40) / (70 × 30) = 0.38. This means that the odds that a major arrhythmia occurred was lower for the patients who were taking a beta-blocker compared to those who were not.

**(Choice E)** Relative risk (RR) is the risk of disease in the exposed divided by the risk of disease in the non-exposed; it is calculated using the formula: RR = [a/(a+b)] / [c/(c+d)] = (20/90) / (30/70). RR and OR are frequently used measures of association that are commonly assessed. However, only OR should be used in a case-control study because such studies do not follow patients over time to calculate a risk. The formula for OR (= ad/bc) is the same in case-control studies (although its mathematical derivation is slightly different).

**Educational objective:**
The odds ratio (OR) is a measure of association between an exposure and an outcome. In a standard contingency table, OR = ad / bc.

Time Spent: 11 seconds



Last updated: [12/15/2015]

A meta-analysis of several trials on the effect of cocoa intake on systolic blood pressure (SBP) revealed the following results:

| Study | Mean SBP in cocoa group − mean SBP in control group (mm Hg) [95% confidence interval] |
|---|---|
| A | -5.2 [-7.3, -3.4] |
| B | 1.6 [-5.3, 10.4] |
| C | -4.2 [-8.1, -2.7] |
| D | -2.9 [-4.1, -1.4] |
| E | -2.8 [-5.2, -1.1] |
| F | 0.8 [0.1, 1.2] |
| G | 1.0 [-1.2, 3.3] |
| **Total** | **-2.2 [-2.7, -1.3]** |

All the trials evaluated the difference in SBP at 2 weeks. Based on the data, which of the following is the most appropriate conclusion?

- A. A higher mean systolic blood pressure was seen in the cocoa groups overall
- B. Cocoa intake should be recommended for blood pressure management
- C. Cocoa intake was associated with a statistically significant decrease in systolic blood pressure

| Study | Mean SBP in cocoa group − mean SBP in control group (mm Hg) [95% confidence interval] |
|---|---|
| A | -5.2 [-7.3, -3.4] |
| B | 1.6 [-5.3, 10.4] |
| C | -4.2 [-8.1, -2.7] |
| D | -2.9 [-4.1, -1.4] |
| E | -2.8 [-5.2, -1.1] |
| F | 0.8 [0.1, 1.2] |
| G | 1.0 [-1.2, 3.3] |
| **Total** | **-2.2 [-2.7, -1.3]** |

All the trials evaluated the difference in SBP at 2 weeks. Based on the data, which of the following is the most appropriate conclusion?

- A. A higher mean systolic blood pressure was seen in the cocoa groups overall
- B. Cocoa intake should be recommended for blood pressure management
- C. Cocoa intake was associated with a statistically significant decrease in systolic blood pressure
- D. Studies B, F, and G showed a statistically significant increase in systolic blood pressure
- E. There was no statistically significant change in systolic blood pressure overall

| Study | (mm Hg) [95% confidence interval] |
|---|---|
| A | -5.2 [-7.3, -3.4] |
| B | 1.6 [-5.3, 10.4] |
| C | -4.2 [-8.1, -2.7] |
| D | -2.9 [-4.1, -1.4] |
| E | -2.8 [-5.2, -1.1] |
| F | 0.8 [0.1, 1.2] |
| G | 1.0 [-1.2, 3.3] |
| **Total** | **-2.2 [-2.7, -1.3]** |

All the trials evaluated the difference in SBP at 2 weeks. Based on the data, which of the following is the most appropriate conclusion?

- A. A higher mean systolic blood pressure was seen in the cocoa groups overall [4%]
- B. Cocoa intake should be recommended for blood pressure management [3%]
- ✔ C. Cocoa intake was associated with a statistically significant decrease in systolic blood pressure [64%]
- D. Studies B, F, and G showed a statistically significant increase in systolic blood pressure [9%]
- E. There was no statistically significant change in systolic blood pressure overall [21%]

**Explanation:**

A **meta-analysis** groups the results of several trials (ideally high-quality randomized controlled trials) to increase statistical power and provide an overall estimate of the effect of an exposure (eg, cocoa intake) on an outcome (eg, systolic blood pressure [SBP]). The individual results of the trials are assessed and weighted to obtain a **pooled effect estimate**. In this example, the results of the meta-analysis showed that there was an overall change in mean SBP of -2.2 mm Hg (ie, decrease in SBP) in the cocoa intake group compared to the control group. The corresponding 95% confidence interval (CI), which is a measure of statistical significance, was -2.7 to -1.3 mm Hg. This range does not include the null value (0 mm Hg in this case, as explained below); therefore, the result is statistically significant. In summary, cocoa intake was associated with a statistically significant decrease in SBP.

A CI is a range that can be interpreted as follows: If the study were repeated 100 times, the results obtained would lie within that range in 95 out of the 100 times. All CIs have a null value. In this example, the null value would correspond to 0 mm Hg (no difference between the cocoa intake and control groups). If the CI crosses the null value (eg, studies B and G), then there is no statistically significant difference between the groups **(Choice D)**. In this example, because the 95% CI of [-2.7, -1.3] is entirely negative (does not cross 0), the result is considered statistically significant.

**(Choices A and E)** The overall mean change in SBP was negative at -2.2 mm Hg, reflecting a lower SBP in the cocoa group compared to the control group. The change was statistically significant.

**(Choice B)** Further research should be conducted before recommending cocoa intake to manage SBP. Although there was a statistically significant decrease in SBP in the cocoa intake group, the clinical significance may be limited as the absolute SBP decrease is only about 2 mm Hg. Furthermore, the trials measured SBP at 2

in the cocoa intake group compared to the control group. The corresponding 95% confidence interval (CI), which is a measure of statistical significance, was -2.7 to -1.3 mm Hg. This range does not include the null value (0 mm Hg in this case, as explained below); therefore, the result is statistically significant. In summary, cocoa intake was associated with a statistically significant decrease in SBP.

A CI is a range that can be interpreted as follows: If the study were repeated 100 times, the results obtained would lie within that range in 95 out of the 100 times. All CIs have a null value. In this example, the null value would correspond to 0 mm Hg (no difference between the cocoa intake and control groups). If the CI crosses the null value (eg, studies B and G), then there is no statistically significant difference between the groups **(Choice D)**. In this example, because the 95% CI of [-2.7, -1.3] is entirely negative (does not cross 0), the result is considered statistically significant.

**(Choices A and E)** The overall mean change in SBP was negative at -2.2 mm Hg, reflecting a lower SBP in the cocoa group compared to the control group. The change was statistically significant.

**(Choice B)** Further research should be conducted before recommending cocoa intake to manage SBP. Although there was a statistically significant decrease in SBP in the cocoa intake group, the clinical significance may be limited as the absolute SBP decrease is only about 2 mm Hg. Furthermore, the trials measured SBP at 2 weeks, so it is possible that the effect of cocoa intake on SBP was short-lived. Finally, no information is provided regarding adverse effects of cocoa intake.

**Educational objective:**
A meta-analysis groups results of several trials to increase statistical power and provide an overall pooled effect estimate.

Time Spent: 28 seconds



Last updated: [9/1/2015]

Researchers are interested in the association between colorectal carcinoma and nonsteroidal anti-inflammatory drug use. They first interview a group of patients with biopsy-proven colorectal carcinoma and then interview a group consisting of the patients' neighbors who are of similar age and race. The analysis is based on comparisons of the results of pairs of individuals with similar characteristics (one from each of the 2 groups). This design technique best helps address which of the following potential problems with this study?

- A. Ascertainment bias
- B. Confounding
- C. Observer bias
- D. Recall bias
- E. Selection bias

Researchers are interested in the association between colorectal carcinoma and nonsteroidal anti-inflammatory drug use. They first interview a group of patients with biopsy-proven colorectal carcinoma and then interview a group consisting of the patients' neighbors who are of similar age and race. The analysis is based on comparisons of the results of pairs of individuals with similar characteristics (one from each of the 2 groups). This design technique best helps address which of the following potential problems with this study?

- A. Ascertainment bias [4%]
- ✔ B. Confounding [45%]
- C. Observer bias [3%]
- D. Recall bias [6%]
- E. Selection bias [43%]

**Explanation:**

**Matching** is a method generally used in the design stage of case-control studies to control confounding (ie, when a perceived association between an exposure and an outcome is actually explained by a confounding variable associated with both the exposure and the outcome). The initial step in matching involves selecting variables that could be confounders (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a **similar distribution** in accordance with the variables.

In this scenario, the "cases" (patients with colorectal cancer) were matched with neighborhood "controls" of similar age and race. Selecting neighbors as controls

**Explanation:**

**Matching** is a method generally used in the design stage of case-control studies to control confounding (ie, when a perceived association between an exposure and an outcome is actually explained by a confounding variable associated with both the exposure and the outcome). The initial step in matching involves selecting variables that could be confounders (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a **similar distribution** in accordance with the variables.

In this scenario, the "cases" (patients with colorectal cancer) were matched with neighborhood "controls" of similar age and race. Selecting neighbors as controls has another advantage of matching the cases to controls by variables that are difficult to measure (eg, socioeconomic status, environmental factors). Gender and smoking status are other common confounders.

**(Choices A, C, D, and E)** Observer bias and ascertainment bias result from mislabeling exposed/unexposed or cases/controls. Recall bias could be a limitation of this study as the interviewed participants with colorectal cancer may be more likely to recall certain exposures. Selection bias is a potential problem in this study because the controls selected may not reflect the exposure experience of the general population. However, although these biases may be present, matching best addresses confounding rather than any of these biases.

**Educational objective:**
Matching is used in case-control studies in order to control confounding. Matching variables should always be the potential confounders of the study (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a similar distribution in accordance with the variables.

Plasma homocysteine levels are measured in patients with acute coronary syndrome who are treated at a large community hospital. The mean plasma homocysteine level in this group is determined to be 11.1 µmol/L with a standard deviation of 1.2 µmol/L. In a separate group of patients hospitalized on the general ward in the same hospital, the mean plasma level is 9.5 µmol/L and the standard deviation is 1.3 µmol/L. Which of the following statistical methods should be used to compare the mean homocysteine levels of these 2 groups of patients?

- A. Two-sample t test
- B. Linear regression
- C. Correlation coefficient
- D. Chi-square test
- E. Meta-analysis

Plasma homocysteine levels are measured in patients with acute coronary syndrome who are treated at a large community hospital. The mean plasma homocysteine level in this group is determined to be 11.1 µmol/L with a standard deviation of 1.2 µmol/L. In a separate group of patients hospitalized on the general ward in the same hospital, the mean plasma level is 9.5 µmol/L and the standard deviation is 1.3 µmol/L. Which of the following statistical methods should be used to compare the mean homocysteine levels of these 2 groups of patients?

- ✔ (●) A. Two-sample t test [75%]
- ( ) B. Linear regression [1%]
- ( ) C. Correlation coefficient [5%]
- ( ) D. Chi-square test [17%]
- ( ) E. Meta-analysis [2%]

**Explanation:**

## Two-sample t test

| Null hypothesis | Alternate hypothesis |
|---|---|
| Both samples drawn from the same population ($\mu1 = \mu2$) | Both samples drawn from different populations ($\mu1 \neq \mu2$) |

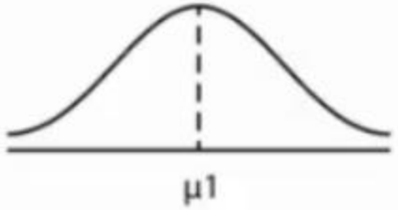
Only 2

**Explanation:**

## Two-sample t test

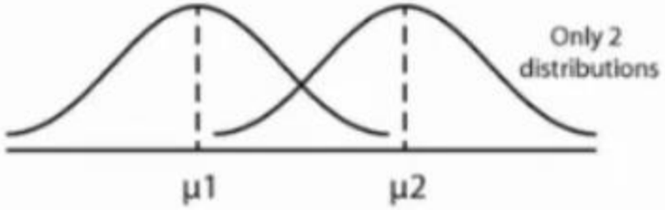
**Null hypothesis**

Both samples drawn from the same population ($\mu 1 = \mu 2$)

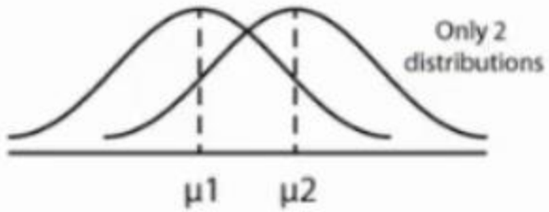



**Alternate hypothesis**

Both samples drawn from different populations ($\mu 1 \neq \mu 2$)

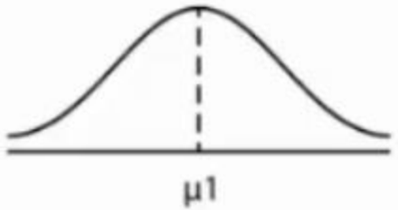



Small apparent difference between the means of the samples
null hypothesis not rejected

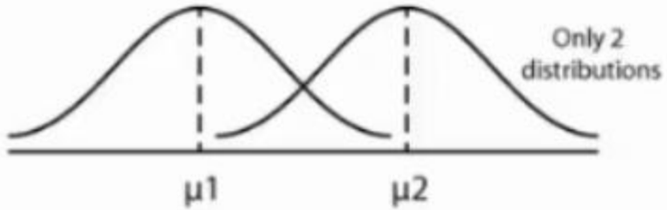



Large apparent differences between the means of the samples
null hypothesis rejected

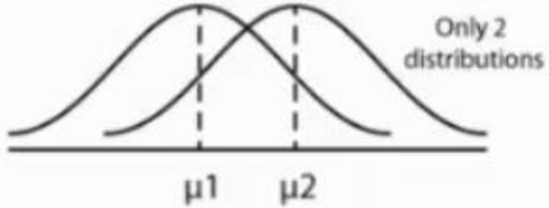
Both samples drawn from the same population ($\mu1 = \mu2$)



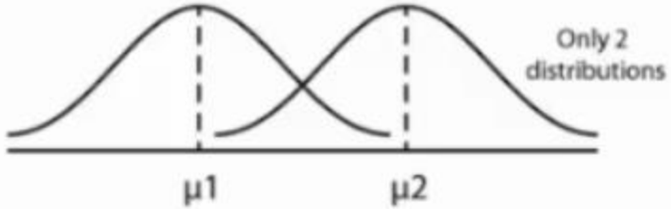

Both samples drawn from different populations ($\mu1 \neq \mu2$)




Small apparent difference between the means of the samples
null hypothesis not rejected




Large apparent differences between the means of the samples
null hypothesis rejected



The two-sample t test is commonly employed to determine if the means of 2 populations are equal. Several statistical approaches can be used, but the basic numerical requirements needed to perform this test are the 2 mean values, the sample variances (eg, standard deviations), and the sample sizes. The t statistic is then calculated, from which the p value can be determined. If $p < 0.05$, the null hypothesis (which assumes that there is no difference between 2 groups) is rejected and the 2 means are assumed to be statistically different.

**(Choice B)** Linear regression is used to model the linear relationship between a dependent variable and an independent variable. For example, linear regression could be used to determine the relationship (described in terms of a trend line) between the number of cigarettes smoked per day and the number of yearly hospitalizations in COPD patients.

**(Choice C)** The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between estrogen levels and breast cancer risk in postmenopausal women. It is different from linear regression in that a single number is reported describing the strength and magnitude of the association.

**(Choice D)** The chi-square test is most appropriate for use with categorical data. It can be used to evaluate whether the expected frequency of an occurrence is consistent with the observed frequency of that occurrence ("goodness of fit"). For instance, a study evaluating Mendelian inheritance of red and green seed colors would use a chi-square test to compare the observed and expected proportions of each seed type.

**(Choice E)** Meta-analysis is an epidemiologic method of analyzing pooled data from

**(Choice B)** Linear regression is used to model the linear relationship between a dependent variable and an independent variable. For example, linear regression could be used to determine the relationship (described in terms of a trend line) between the number of cigarettes smoked per day and the number of yearly hospitalizations in COPD patients.

**(Choice C)** The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between estrogen levels and breast cancer risk in postmenopausal women. It is different from linear regression in that a single number is reported describing the strength and magnitude of the association.

**(Choice D)** The chi-square test is most appropriate for use with categorical data. It can be used to evaluate whether the expected frequency of an occurrence is consistent with the observed frequency of that occurrence ("goodness of fit"). For instance, a study evaluating Mendelian inheritance of red and green seed colors would use a chi-square test to compare the observed and expected proportions of each seed type.

**(Choice E)** Meta-analysis is an epidemiologic method of analyzing pooled data from several studies, thereby increasing the statistical power beyond that of the individual studies.

**Educational objective:**
The two-sample t test is a statistical method commonly employed to compare the means of 2 groups of subjects.

Time Spent: 8 seconds



Last updated: [10/7/2015]

Researchers conduct a prospective study that demonstrates an association between alcohol consumption and transitional bladder carcinoma, with a relative risk (RR) of 1.81 and p-value of 0.03. They then divide the study subjects into 2 groups, smokers and nonsmokers, and again examine the association between alcohol consumption and bladder cancer:

Smokers: RR = 0.95; p = 0.87
Nonsmokers: RR = 1.03; p = 0.96

The discrepancy between the overall results and the stratified results is best explained by which of the following?

- A. Effect modification
- B. Observer bias
- C. Measurement bias
- D. Recall bias
- E. Confounding
- F. Meta-analysis

Researchers conduct a prospective study that demonstrates an association between alcohol consumption and transitional bladder carcinoma, with a relative risk (RR) of 1.81 and p-value of 0.03. They then divide the study subjects into 2 groups, smokers and nonsmokers, and again examine the association between alcohol consumption and bladder cancer:

Smokers: RR = 0.95; p = 0.87
Nonsmokers: RR = 1.03; p = 0.96

The discrepancy between the overall results and the stratified results is best explained by which of the following?

- A. Effect modification [19%]
- B. Observer bias [2%]
- C. Measurement bias [7%]
- D. Recall bias [1%]
- ✔ E. Confounding [67%]
- F. Meta-analysis [3%]

**Explanation:**

Initial crude analysis confounded by smoking status

Bladder cancer

Yes | No

## Explanation:

**Initial crude analysis confounded by smoking status**

| | | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|---|
| Alcohol use | Yes | 23 | 207 |
| | No | 26 | 444 |

RR = 1.81
p-value= 0.03

**Stratified analysis removes confounding effects of smoking**

**Smokers**

| | | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|---|
| Alcohol use | Yes | 19 | 81 |
| | No | 14 | 56 |

RR = 0.95

**Non-smokers**

| | | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|---|
| Alcohol use | Yes | 4 | 126 |
| | No | 12 | 388 |

RR = 1.03

## Initial crude analysis confounded by smoking status

| Alcohol use | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|
| Yes | 23 | 207 |
| No | 26 | 444 |

RR = 1.81
p-value= 0.03

## Stratified analysis removes confounding effects of smoking

**Smokers**

| Alcohol use | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|
| Yes | 19 | 81 |
| No | 14 | 56 |

RR = 0.95
p-value= 0.87

**Non-smokers**

| Alcohol use | Bladder cancer: Yes | Bladder cancer: No |
|---|---|---|
| Yes | 4 | 126 |
| No | 12 | 388 |

RR = 1.03
p-value= 0.96

This scenario depicts a typical example of confounding. Confounding bias occurs when the exposure-disease relationship is muddied by the effect of a confounding variable. Confounders are extraneous factors that correlate with both the exposure and the disease. The initial crude analysis of the study data suggested that alcohol consumption was associated with bladder carcinoma. However, this analysis did not account for the effect of smoking, a potential confounder. Smoking is associated with both bladder carcinoma and significant alcohol consumption (ie, people who drink a lot are more likely to smoke). As a result, smoking is a potential confounder that may explain some or all of the association observed between alcohol consumption and bladder cancer.

Confounding effects can be negated by running separate analyses for the smokers and non-smokers (a technique called stratified analysis). This allows calculation of the true, unconfounded RR that alcohol users have of developing bladder cancer. In the example above, stratified analysis shows that both the smoking and nonsmoking groups have an RR close to 1. Their large p-values further suggest that the small difference in the RR from the null value of 1 is due to chance alone. Thus, the RR of 1.81 found on crude analysis disappears and there is no true association between alcohol consumption and bladder cancer.

**(Choice A)** Effect modification results when an external variable positively or negatively impacts the observed effect of a risk factor on disease status. When this occurs, stratified analysis will reveal a significant difference in risk between the stratified groups. Because there is no significant difference in RR between the smoking and nonsmoking groups (p-values > 0.05), smoking status does not modify the effect of alcohol consumption on bladder cancer. Thus, smoking status is a confounder and not an effect modifier.

**(Choices B and C)** Observer bias and measurement bias distort the strength of

difference in the RR from the null value of 1 is due to chance alone. Thus, the RR of 1.81 found on crude analysis disappears and there is no true association between alcohol consumption and bladder cancer.

**(Choice A)** Effect modification results when an external variable positively or negatively impacts the observed effect of a risk factor on disease status. When this occurs, stratified analysis will reveal a significant difference in risk between the stratified groups. Because there is no significant difference in RR between the smoking and nonsmoking groups (p-values > 0.05), smoking status does not modify the effect of alcohol consumption on bladder cancer. Thus, smoking status is a confounder and not an effect modifier.

**(Choices B and C)** Observer bias and measurement bias distort the strength of association by misclassifying exposed/unexposed and/or diseased/nondiseased subjects. The scenario does not describe any particular issues that could affect the classification process.

**(Choice D)** Recall bias results from inaccurate recall of past exposure by people in the study and applies mostly to retrospective studies such as case-control studies.

**(Choice F)** Meta-analysis refers to compiling data from several studies to increase the power of analysis.

**Educational objective:**
Confounding bias occurs when the exposure-disease relationship is muddied by the effect of an extraneous factor that has correlations with both the exposure and the disease. Confounding bias can result in the false association of an exposure with a disease.

Time Spent: 7 seconds



Last updated: [10/7/2015]

A study to assess spironolactone's efficacy in patients with heart failure is performed. 450 patients receive either spironolactone or placebo for two years. Neither the patients nor physicians are aware of who takes the drug or placebo. The study setup described above is most effective in preventing:

- A. Beta error
- B. Recall bias
- C. Observer bias
- D. Effect modification
- E. Selection bias

A study to assess spironolactone's efficacy in patients with heart failure is performed. 450 patients receive either spironolactone or placebo for two years. Neither the patients nor physicians are aware of who takes the drug or placebo. The study setup described above is most effective in preventing:

- A. Beta error [3%]
- B. Recall bias [2%]
- ✔ C. Observer bias [74%]
- D. Effect modification [8%]
- E. Selection bias [12%]

**Explanation:**

Blinding technique is commonly used in clinical trials. The blinding can involve patients exclusively or both patients and physicians (double blinding). The main purpose of blinding is to prevent patient or researcher expectancy from interfering with the determination of an outcome. For example, a researcher's belief in a positive outcome in treated patients can potentially result in observer bias.

**(Choice A)** Beta error refers to a conclusion that there is no difference between the groups studied when a difference truly exists. Beta error is a random error, not a systematic error (i.e., bias).

**(Choice B)** Recall bias results from the inaccurate recall of past exposure by subjects. It applies mostly to case-control studies.

**(Choice D)** Effect modification is not a bias and should not be controlled

- [x] C. Observer bias [74%]
- [ ] D. Effect modification [8%]
- [ ] E. Selection bias [12%]

**Explanation:**

Blinding technique is commonly used in clinical trials. The blinding can involve patients exclusively or both patients and physicians (double blinding). The main purpose of blinding is to prevent patient or researcher expectancy from interfering with the determination of an outcome. For example, a researcher's belief in a positive outcome in treated patients can potentially result in observer bias.

**(Choice A)** Beta error refers to a conclusion that there is no difference between the groups studied when a difference truly exists. Beta error is a random error, not a systematic error (i.e., bias).

**(Choice B)** Recall bias results from the inaccurate recall of past exposure by subjects. It applies mostly to case-control studies.

**(Choice D)** Effect modification is not a bias and should not be controlled.

**(Choice E)** Selection bias results from the manner in which people are selected for the study, or from the selective losses from follow-up.

**Educational Objective:**
The main purpose of blinding is to prevent patient or researcher expectancy from interfering with an outcome.

Time Spent: 6 seconds



Last updated: [8/19/2015]

A suburban hospital's yearly surgical infection statistics are shown below.

| Surgical infection | Fatal | % of fatal | Non-fatal | % of non-fatal |
|---|---|---|---|---|
| *Cl. difficile* | 2 | 3 | 14 | 6 |
| *S. aureus*, methicillin-sensitive | 10 | 13 | 25 | 11 |
| *S. aureus*, methicillin-resistant | 40 | 53 | 70 | 32 |
| *E. coli* | 10 | 13 | 50 | 23 |
| *Cl. perfringens* | 1 | 1 | 1 | < 1 |
| *S. epidermidis* | 2 | 3 | 30 | 14 |
| Others | 10 | 13 | 30 | 14 |
| Overall | 75 | 100 | 220 | 100 |

What is the case-fatality rate for methicillin-resistant *S. aureus* surgical infections in this hospital?

- A. 40/70
- B. 40/110
- C. 40/75
- D. 53/100
- E. 70/220

A suburban hospital's yearly surgical infection statistics are shown below.

| Surgical infection | Fatal | % of fatal | Non-fatal | % of non-fatal |
|---|---|---|---|---|
| *Cl. difficile* | 2 | 3 | 14 | 6 |
| *S. aureus*, methicillin-sensitive | 10 | 13 | 25 | 11 |
| *S. aureus*, methicillin-resistant | 40 | 53 | 70 | 32 |
| *E. coli* | 10 | 13 | 50 | 23 |
| *Cl. perfringens* | 1 | 1 | 1 | < 1 |
| *S. epidermidis* | 2 | 3 | 30 | 14 |
| Others | 10 | 13 | 30 | 14 |
| Overall | 75 | 100 | 220 | 100 |

What is the case-fatality rate for methicillin-resistant *S. aureus* surgical infections in this hospital?

- A. 40/70 [7%]
- ✔ B. 40/110 [72%]
- C. 40/75 [13%]
- D. 53/100 [6%]
- E. 70/220 [1%]

**Explanation:**

You may encounter USMLE scenarios where you are asked to interpret epidemiological graphs or tables and perform some basic calculations. Case-fatality rate is calculated by dividing the number of fatal cases by the total number of people with the disease. In this scenario, 40 fatal and 70 non-fatal cases of MRSA infection

| | | | | |
|---|---|---|---|---|
| Others | 10 | 13 | 30 | 14 |
| Overall | 75 | 100 | 220 | 100 |

What is the case-fatality rate for methicillin-resistant *S. aureus* surgical infections in this hospital?

- A. 40/70 [7%]
- ✔ B. 40/110 [72%]
- C. 40/75 [13%]
- D. 53/100 [6%]
- E. 70/220 [1%]

**Explanation:**

You may encounter USMLE scenarios where you are asked to interpret epidemiological graphs or tables and perform some basic calculations. Case-fatality rate is calculated by dividing the number of fatal cases by the total number of people with the disease. In this scenario, 40 fatal and 70 non-fatal cases of MRSA infection are reported. Therefore, case-fatality rate is 40/(40+70) = 40/110.

**(Choices C and D)** These choices describe the proportion of MRSA deaths out of all surgical infection deaths (40/75).

**Educational Objective:**
Case-fatality rate is calculated by dividing the number of fatal cases by the total number of people with the disease.

Time Spent: 11 seconds



Last updated: [11/6/2015]

A study evaluated the role of angiotensin-converting enzyme (ACE) inhibitors in the prevention of coronary events in patients with diabetes. During 5 years of follow-up, 120 out of 400 diabetic patients who had taken an ACE inhibitor developed an acute coronary event. Over the same time, 100 out of 300 diabetic patients who had not taken an ACE inhibitor experienced coronary events. What was the relative risk of developing a coronary event in diabetic patients who were taking ACE inhibitors compared to diabetic patients who were not taking ACE inhibitors?

- A. (120x100)/(280x200)
- B. (120x200)/(100x280)
- C. (120x280)/(100x200)
- D. (120/220)/(280/480)
- E. (120/400)/(100/300)

A study evaluated the role of angiotensin-converting enzyme (ACE) inhibitors in the prevention of coronary events in patients with diabetes. During 5 years of follow-up, 120 out of 400 diabetic patients who had taken an ACE inhibitor developed an acute coronary event. Over the same time, 100 out of 300 diabetic patients who had not taken an ACE inhibitor experienced coronary events. What was the relative risk of developing a coronary event in diabetic patients who were taking ACE inhibitors compared to diabetic patients who were not taking ACE inhibitors?

- A. (120x100)/(280x200) [3%]
- B. (120x200)/(100x280) [11%]
- C. (120x280)/(100x200) [4%]
- D. (120/220)/(280/480) [4%]
- ✔ E. (120/400)/(100/300) [77%]

**Explanation:**

It is important to know how to calculate some basic measures of association when given raw data. The first step is to organize the data into a 2×2 (contingency) table as shown below. The standard format includes the exposures in the rows (with "exposure present" listed above "exposure absent") and the outcomes in the columns (with "outcome present" listed before "outcome absent").

| | Coronary event | No coronary event | |
|---|---|---|---|
| ACE inhibitor | 120 (a) | 280 (b) | 400 |

**Explanation:**

It is important to know how to calculate some basic measures of association when given raw data. The first step is to organize the data into a 2×2 (contingency) table as shown below. The standard format includes the exposures in the rows (with "exposure present" listed above "exposure absent") and the outcomes in the columns (with "outcome present" listed before "outcome absent").

| | Coronary event | No coronary event | |
|---|---|---|---|
| ACE inhibitor | 120 (a) | 280 (b) | 400 |
| No ACE inhibitor | 100 (c) | 200 (d) | 300 |
| | 220 | 480 | 700 |

The relative risk (RR) represents the risk of an outcome (eg, coronary event) in the exposed (eg, angiotensin-converting enzyme [ACE] inhibitor) divided by the risk of that outcome in the unexposed (eg, no ACE inhibitor). For cohort studies, it is calculated using the following formula:

$$RR = [a/(a+b)] / [c/(c+d)] = (120/400)/(100/300) = 0.90$$

Understanding the values calculated (rather than memorizing several formulas) would prevent errors that could occur if the standard format for the 2×2 table is not followed. By definition, RR = risk among the exposed/risk among the unexposed. In this example, 400 diabetic patients were taking ACE inhibitors (exposed) and 120 of those had a coronary event. Therefore, the risk among the exposed is 120/400.

| | | | |
|---|---|---|---|
| | 220 | 480 | 700 |

The relative risk (RR) represents the risk of an outcome (eg, coronary event) in the exposed (eg, angiotensin-converting enzyme [ACE] inhibitor) divided by the risk of that outcome in the unexposed (eg, no ACE inhibitor). For cohort studies, it is calculated using the following formula:

RR = [a/(a+b)] / [c/(c+d)] = (120/400)/(100/300) = 0.90

Understanding the values calculated (rather than memorizing several formulas) would prevent errors that could occur if the standard format for the 2×2 table is not followed. By definition, RR = risk among the exposed/risk among the unexposed. In this example, 400 diabetic patients were taking ACE inhibitors (exposed) and 120 of those had a coronary event. Therefore, the risk among the exposed is 120/400. Similarly, 300 diabetic patients were not taking ACE inhibitors (unexposed) and 100 of those had a coronary event. Therefore, the risk among the unexposed is 100/300. The RR is the ratio of those 2 values: RR = (120/400)/(100/300).

**(Choice B)** The odds ratio (OR) is calculated using the following formula: OR = ad/bc. The OR is often calculated in case-control studies where RR cannot be obtained because such studies do not follow patients over time to determine the incidence of a new outcome.

**Educational objective:**
The relative risk (RR) represents the risk of an outcome in the exposed divided by the risk of that outcome in the unexposed. Applying the correct formula for RR calculations depends on the proper formatting of a 2×2 (contingency) table.

Time Spent: 11 seconds



Last updated: [8/20/2015]

A new biomarker has been identified that allows for the early detection of invasive gastric carcinoma. It has a sensitivity of 89% and a specificity of 85% when compared to endoscopy with multiple biopsies. The test is used in 2 groups: a population in the United States, where 5 out of 100,000 people have gastric cancer, and a population in China, where 100 out of 100,000 people have been diagnosed with the disease. Which of the following is the most accurate statement concerning this new test?

- A. Negative predictive value of the test is lower in the American population
- B. Positive predictive value of the test is higher in the Chinese population
- C. Sensitivity of the test is higher in the American population
- D. Specificity of the test is higher in the Chinese population
- E. The test is not reliable in the American population

A new biomarker has been identified that allows for the early detection of invasive gastric carcinoma. It has a sensitivity of 89% and a specificity of 85% when compared to endoscopy with multiple biopsies. The test is used in 2 groups: a population in the United States, where 5 out of 100,000 people have gastric cancer, and a population in China, where 100 out of 100,000 people have been diagnosed with the disease. Which of the following is the most accurate statement concerning this new test?

- A. Negative predictive value of the test is lower in the American population [5%]
- ✔ B. Positive predictive value of the test is higher in the Chinese population [83%]
- C. Sensitivity of the test is higher in the American population [4%]
- D. Specificity of the test is higher in the Chinese population [4%]
- E. The test is not reliable in the American population [4%]

**Explanation:**

| | Positive condition | Negative condition | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |

**Explanation:**

| | Positive condition | Negative condition | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |
| | Sensitivity = TP/(TP+FN) | Specificity = TN/(TN+FP) | |

**FN** = false negative; **FP** = false positive; **TN** = true negative; **TP** = true positive; **NPV** = negative predictive value; **PPV** = positive predictive value.



The **positive predictive value** (PPV) is defined as the proportion of individuals with positive test results who actually have the disease. This concept can be understood in terms of probability: given that a patient has a positive test result, what is the likelihood that he or she actually has the disease? PPV depends not only on the sensitivity and specificity of a test, but also on the **prevalence** of the disease in the population being tested. The more common the disease in the population, the more likely it is that a patient with a positive test result actually has the disease.

If the disease is relatively common (eg, gastric cancer in China), there is a higher probability that a patient who tests positive actually has the disease (ie, more likely to be a true positive). However, if the disease is relatively uncommon (eg, gastric cancer in the United States), a patient who tests positive will have a lower probability of actually having the disease (ie, more likely to be a false positive).

The **positive predictive value** (PPV) is defined as the proportion of individuals with positive test results who actually have the disease. This concept can be understood in terms of probability: given that a patient has a positive test result, what is the likelihood that he or she actually has the disease? PPV depends not only on the sensitivity and specificity of a test, but also on the **prevalence** of the disease in the population being tested. The more common the disease in the population, the more likely it is that a patient with a positive test result actually has the disease.

If the disease is relatively common (eg, gastric cancer in China), there is a higher probability that a patient who tests positive actually has the disease (ie, more likely to be a true positive). However, if the disease is relatively uncommon (eg, gastric cancer in the United States), a patient who tests positive will have a lower probability of actually having the disease (ie, more likely to be a false positive).

**(Choice A)** Like PPV, the negative predictive value (NPV) depends on disease prevalence. However, NPV is inversely associated with the prevalence: NPV increases as the disease prevalence decreases. This is because the probability of a true negative result is higher in a population with low disease prevalence (eg, gastric cancer in the United States).

**(Choices C and D)** Sensitivity and specificity are intrinsic test parameters that do not depend on the prevalence of disease in the tested population.

**(Choice E)** Reliability is a measure of the consistency (reproducibility) of a diagnostic test result (ie, if the test is taken multiple times, will the results be the same?). No information is provided to suggest that the gastric cancer biomarker is unreliable.

**Educational objective:**

sensitivity and specificity of a test, but also on the **prevalence** of the disease in the population being tested. The more common the disease in the population, the more likely it is that a patient with a positive test result actually has the disease.

If the disease is relatively common (eg, gastric cancer in China), there is a higher probability that a patient who tests positive actually has the disease (ie, more likely to be a true positive). However, if the disease is relatively uncommon (eg, gastric cancer in the United States), a patient who tests positive will have a lower probability of actually having the disease (ie, more likely to be a false positive).

**(Choice A)** Like PPV, the negative predictive value (NPV) depends on disease prevalence. However, NPV is inversely associated with the prevalence: NPV increases as the disease prevalence decreases. This is because the probability of a true negative result is higher in a population with low disease prevalence (eg, gastric cancer in the United States).

**(Choices C and D)** Sensitivity and specificity are intrinsic test parameters that do not depend on the prevalence of disease in the tested population.

**(Choice E)** Reliability is a measure of the consistency (reproducibility) of a diagnostic test result (ie, if the test is taken multiple times, will the results be the same?). No information is provided to suggest that the gastric cancer biomarker is unreliable.

**Educational objective:**
Positive and negative predictive values depend on disease prevalence in the tested population. The sensitivity and specificity of a diagnostic test are not affected by disease prevalence.

Time Spent: 14 seconds



Last updated: [9/19/2015]

Researchers studying the effects of hormone replacement therapy (HRT) on the risk of myocardial infarction (MI) among postmenopausal women calculate the relative risk (RR) of MI to be 1.30 ($p = 0.07$) among women who are taking HRT compared to those who are not. The researchers conclude that there is no statistically significant increased risk of MI with HRT (based on a cutoff of $\alpha = 0.05$). Subsequently, the results of a meta-analysis determine that there actually is an increased risk of MI, with an overall RR = 1.32 ($p = 0.03$) among postmenopausal women who are taking HRT compared to those who are not. Which of the following was the most likely problem in the first study?

- A. Berkson's bias
- B. Placebo effect
- C. Poor blinding
- D. Researcher expectancy
- E. Sample size

Researchers studying the effects of hormone replacement therapy (HRT) on the risk of myocardial infarction (MI) among postmenopausal women calculate the relative risk (RR) of MI to be 1.30 (p = 0.07) among women who are taking HRT compared to those who are not. The researchers conclude that there is no statistically significant increased risk of MI with HRT (based on a cutoff of α = 0.05). Subsequently, the results of a meta-analysis determine that there actually is an increased risk of MI, with an overall RR = 1.32 (p = 0.03) among postmenopausal women who are taking HRT compared to those who are not. Which of the following was the most likely problem in the first study?

- A. Berkson's bias [9%]
- B. Placebo effect [2%]
- C. Poor blinding [5%]
- D. Researcher expectancy [13%]
- ✔ E. Sample size [71%]

**Explanation:**

**Type I (α) and type II (β) errors**

| | True status | |
|---|---|---|
| | **There is a true difference** (ie, $H_0$ is false) | **There is NO true difference** (ie, $H_0$ is true) |

**Explanation:**

**Type I (α) and type II (β) errors**

| | | True status | |
|---|---|---|---|
| | | **There is a true difference** (ie, $H_0$ is false) | **There is NO true difference** (ie, $H_0$ is true) |
| **Study result** | **Difference calculated as statistically significant** (ie, reject $H_0$) | **Correctly** conclude there is a difference | **Type I (α) error** (Falsely conclude there is a difference) |
| | **Difference calculated as NOT statistically significant** (ie, fail to reject $H_0$) | **Type II (β) error** (Falsely conclude there is NO difference) | **Correctly** conclude there is NO difference |

$H_0$ = null hypothesis of no difference.



The first study and the meta-analysis (combined results of several studies) had similar outcomes in terms of relative risk (RR); however, they arrived at different conclusions because the p-value reached statistical significance ($p < \alpha = 0.05$) in the meta-analysis but not in the first study. The most probable reason behind this is that the larger meta-analysis reflects the true status (increased risk of myocardial

The first study and the meta-analysis (combined results of several studies) had similar outcomes in terms of relative risk (RR); however, they arrived at different conclusions because the p-value reached statistical significance ($p < \alpha = 0.05$) in the meta-analysis but not in the first study. The most probable reason behind this is that the **larger** meta-analysis reflects the true status (increased risk of myocardial infarction [MI] with hormone replacement therapy [HRT]), whereas the first study result represented a **type II (β) error** (**falsely** concluded there was **no** increased risk of MI with HRT).

The probability of a type II (β) error is related to how much **power** a study has to detect a difference when a difference actually exists (power = 1 – β). **Sample size** and power are related in that studies with a larger sample size have greater power to detect differences if these exist. The small sample size of the first study made it underpowered to detect a difference in outcome between patients who were treated with HRT and those who were not.

The sample size did not change the value of α (typically set at 0.05). However, because the first study was underpowered, the p-value obtained was not statistically significant ($p = 0.07 > \alpha = 0.05$), leading to a false conclusion of no increased risk of MI with HRT.

**(Choice A)** Berkson's bias refers to selection bias that can be created by selecting hospitalized patients as the control group.

**(Choice B)** Placebo effect refers to patients' expectations affecting an outcome. However, in this case, the control group is taking neither HRT nor a placebo medication.

**(Choices C and D)** Given that the RR obtained in the first study was similar to the one obtained in the meta-analysis (1.32 versus 1.30), it is unlikely that design flaws

result represented a **type II (β) error** (**falsely** concluded there was **no** increased risk of MI with HRT).

The probability of a type II (β) error is related to how much **power** a study has to detect a difference when a difference actually exists (power = 1 – β). **Sample size** and power are related in that studies with a larger sample size have greater power to detect differences if these exist. The small sample size of the first study made it underpowered to detect a difference in outcome between patients who were treated with HRT and those who were not.

The sample size did not change the value of α (typically set at 0.05). However, because the first study was underpowered, the p-value obtained was not statistically significant ($p = 0.07 > \alpha = 0.05$), leading to a false conclusion of no increased risk of MI with HRT.

**(Choice A)** Berkson's bias refers to selection bias that can be created by selecting hospitalized patients as the control group.

**(Choice B)** Placebo effect refers to patients' expectations affecting an outcome. However, in this case, the control group is taking neither HRT nor a placebo medication.

**(Choices C and D)** Given that the RR obtained in the first study was similar to the one obtained in the meta-analysis (1.32 versus 1.30), it is unlikely that design flaws (eg, poor blinding, researcher expectancy) were present in the study.

**Educational objective:**
A study's power increases as its sample size increases. Therefore, the larger the sample, the greater the ability of a study to detect a difference when one truly exists.

Time Spent: 11 seconds



Last updated: [12/14/2015]

A residency program organized a barbecue at the beginning of the academic year to welcome the new class of first-year residents. The total number of attendees was 100, including faculty, administrative staff, and residents. Hamburgers, hotdogs, and potato salad were served. The following day, 28 of the attendees had diarrhea and vomiting. All of the attendees were questioned about what they had eaten, and the following table was obtained:

| Food item or combination of items | Number of attendees who ate food item or combination of items | Number of attendees who developed diarrhea and vomiting |
|---|---|---|
| Hamburgers only | 15 | 2 |
| Hotdogs only | 12 | 1 |
| Potato salad only | 10 | 3 |
| Hamburgers and potato salad | 25 | 5 |
| Hotdogs and potato salad | 8 | 3 |
| Hamburgers, hotdogs, and potato salad | 30 | 14 |

Which of the following best describes the attack rate among all of the attendees who had potato salad?

| Items | Combination of items | diarrhea and vomiting |
|---|---|---|
| Hamburgers only | 15 | 2 |
| Hotdogs only | 12 | 1 |
| Potato salad only | 10 | 3 |
| Hamburgers and potato salad | 25 | 5 |
| Hotdogs and potato salad | 8 | 3 |
| Hamburgers, hotdogs, and potato salad | 30 | 14 |

Which of the following best describes the attack rate among all of the attendees who had potato salad?

- A. 8%
- B. 11%
- C. 13%
- D. 20%
- E. 28%
- F. 30%
- G. 34%
- H. 47%

| Items | Combination of Items | Diarrhea and Vomiting |
|---|---|---|
| Hamburgers only | 15 | 2 |
| Hotdogs only | 12 | 1 |
| Potato salad only | 10 | 3 |
| Hamburgers and potato salad | 25 | 5 |
| Hotdogs and potato salad | 8 | 3 |
| Hamburgers, hotdogs, and potato salad | 30 | 14 |

Which of the following best describes the attack rate among all of the attendees who had potato salad?

- A. 8% [1%]
- B. 11% [2%]
- C. 13% [2%]
- D. 20% [3%]
- E. 28% [6%]
- F. 30% [17%]
- ✔ G. 34% [67%]
- H. 47% [3%]

**Explanation:**

There appears to be an **outbreak** as almost a third of the attendees (28/100) developed diarrhea and vomiting; this is clearly a larger number than expected. The **attack rate** is the ratio of the number of individuals who become ill divided by the number of individuals who are at risk of contracting that illness; it is often used in outbreak investigations. The attack rate for potato salad corresponds to the proportion of attendees who became ill out of the total number of attendees who had potato salad (alone or in combination with other food items).

Total number of attendees who had potato salad
= Number of attendees who had potato salad only + number of attendees who had hamburgers and potato salad + number of attendees who had hotdogs and potato salad + number of attendees who had hamburgers, hotdogs, and potato salad
= 10 + 25 + 8 + 30
= 73

Among those, the number who became ill was 25 (3 out of the 10 who had potato salad only + 5 out of the 25 who had hamburgers and potato salad + 3 out of the 8 who had hotdogs and potato salad + 14 out of the 30 who had hamburgers, hot dogs, and potato salad = 25). Therefore, the attack rate for potato salad was: (25/73) × 100 = 34%.

**(Choice B)** There were 27 attendees who did not eat potato salad: the 15 who had hamburgers only and the 12 who had hotdogs only. Of those, a total of 3 (2 out of the 15 who ate hamburgers only + 1 out of the 12 who ate hotdogs only = 3) became ill. Therefore, the attack rate among those who did not eat potato salad was: (3/27) × 100 = 11%. Despite the fact that these individuals only ate hotdogs and hamburgers, a percentage of them still became ill, probably as a result of some

**(Choice B)** There were 27 attendees who did not eat potato salad: the 15 who had hamburgers only and the 12 who had hotdogs only. Of those, a total of 3 (2 out of the 15 who ate hamburgers only + 1 out of the 12 who ate hotdogs only = 3) became ill. Therefore, the attack rate among those who did not eat potato salad was: (3/27) × 100 = 11%. Despite the fact that these individuals only ate hotdogs and hamburgers, a percentage of them still became ill, probably as a result of some contamination of these food items with the potato salad.

For the sake of completion, the attack rates for the other individual food items or combinations of food items are shown in the following table (they do not need to be calculated to arrive at the correct answer):

| | Food item or combination of items | Number of attendees who ate food item or combination of items | Number of attendees who developed diarrhea and vomiting | Attack rate |
|---|---|---|---|---|
| Did not have potato salad = 27 (attack rate 11%) | Hamburgers only | 15 | 2 | (2/15) × 100 = 13% **(Choice C)** |
| | Hotdogs only | 12 | 1 | (1/12) × 100 = 8% **(Choice A)** |
| | Potato salad only | 10 | 3 | (3/10) × 100 = 30% **(Choice E)** |

| | | | | |
|---|---|---|---|---|
| Did not have potato salad = 27 (attack rate 11%) | Hamburgers only | 15 | 2 | (2/15) × 100 = 13% **(Choice C)** |
| | Hotdogs only | 12 | 1 | (1/12) × 100 = 8% **(Choice A)** |
| Had potato salad = 73 (attack rate 34%) | Potato salad only | 10 | 3 | (3/10) × 100 = 30% **(Choice F)** |
| | Hamburgers and potato salad | 25 | 5 | (5/25) × 100 = 20% **(Choice D)** |
| | Hotdogs and potato salad | 8 | 3 | (3/8) × 100 = 38% |
| | Hamburgers, hotdogs, and potato salad | 30 | 14 | (14/30) × 100 = 47% **(Choice H)** |
| Overall | | 100 | 28 | 28/100 × 100 = 28% **(Choice E)** |

The food item responsible for an outbreak often (though not always) has the largest attack rate and the largest difference in attack rates between those who consumed

| | | | | |
|---|---|---|---|---|
| | only | | | **(Choice F)** |
| Had potato salad = 73 (attack rate 34%) | Hamburgers and potato salad | 25 | 5 | (5/25) × 100 = 20% **(Choice D)** |
| | Hotdogs and potato salad | 8 | 3 | (3/8) × 100 = 38% |
| | Hamburgers, hotdogs, and potato salad | 30 | 14 | (14/30) × 100 = 47% **(Choice H)** |
| Overall | | 100 | 28 | 28/100 × 100 = 28% **(Choice E)** |

The food item responsible for an outbreak often (though not always) has the largest attack rate and the largest difference in attack rates between those who consumed the item and those who did not. In this example, as calculated above, the difference in attack rate between those who had potato salad and those who did not was: 34% − 11% = 23%. As seen in the exhibit, the differences in attack rates were smaller for hamburgers (7%) and hotdogs (16%). It is possible that potato salad is the source of the outbreak.

**Educational objective:**
The attack rate is the ratio of the number of people who contract an illness divided by the number of people who are at risk of contracting that illness.

Time Spent: 29 seconds



Last updated: [1/18/2016]

A 62-year-old man with diabetes, hypertension, and hyperlipidemia comes to the emergency department of an academic medical center with chest pain, nausea, vomiting, and diaphoresis. An electrocardiogram demonstrates ST elevation in the anterior leads, and cardiac enzymes are markedly elevated. Investigators at the center are designing a randomized control trial to test the hypothesis that drug B will decrease the mortality associated with acute ST-elevation myocardial infarction compared to standard of care. To ensure that investigators will not miss a difference between drug B and standard of care (if a difference truly exists), which of the following would they want to maximize?

- A. α
- B. β
- C. Type I error
- D. Type II error
- E. 1 – β

A 62-year-old man with diabetes, hypertension, and hyperlipidemia comes to the emergency department of an academic medical center with chest pain, nausea, vomiting, and diaphoresis. An electrocardiogram demonstrates ST elevation in the anterior leads, and cardiac enzymes are markedly elevated. Investigators at the center are designing a randomized control trial to test the hypothesis that drug B will decrease the mortality associated with acute ST-elevation myocardial infarction compared to standard of care. To ensure that investigators will not miss a difference between drug B and standard of care (if a difference truly exists), which of the following would they want to maximize?

- A. α [6%]
- B. β [10%]
- C. Type I error [6%]
- D. Type II error [6%]
- ✔ E. 1 – β [72%]

**Explanation:**

Statistical **power**, (1 – β), represents a study's ability to detect a difference when one exists. It is the probability of rejecting the null hypothesis when it is truly false - ie, the probability of finding a true relationship. Power depends on sample size and the difference in outcome between the groups being tested. In this study, the researchers want to detect a difference between drug B and standard of care if one exists; they want to maximize power.

β is the probability of committing a type II error **(Choices B and D)**. Type II error

**Explanation:**

Statistical **power**, (1 – β), represents a study's ability to detect a difference when one exists. It is the probability of rejecting the null hypothesis when it is truly false - ie, the probability of finding a true relationship. Power depends on sample size and the difference in outcome between the groups being tested. In this study, the researchers want to detect a difference between drug B and standard of care if one exists; they want to maximize power.

β is the probability of committing a type II error **(Choices B and D)**. Type II error occurs when researchers fail to reject the null hypothesis when it is truly false. This causes investigators to miss true relationships. An example of a type II error would be a study finding that aspirin does not affect platelet function when, in fact, it does. Therefore, if β is set at 0.2, the power will be (1 – β) = 80%; there will be an 80% chance of rejecting the null hypothesis when it is truly false.

**(Choices A and C)** Type I error occurs when researchers reject the null hypothesis when the null hypothesis is really true. That is, a study finds a statistically significant difference between 2 groups when one does not truly exist. An example would be if a study concluded that candy improves heart failure mortality when, in fact, it does not. α is the maximum probability of making a type I error that a researcher is willing to accept. Generally, α is compared to the p-value, the probability of observing a given result (or more extreme) due to chance alone assuming the null hypothesis is true (eg, if there is no real difference between the groups). The value of α is typically set at 0.05, meaning that researchers are willing to accept up to a 5% chance of making a type I error. In such a scenario, if $p < 0.05$, the result is said to be statistically significant.

**Educational objective:**

ie, the probability of finding a true relationship. Power depends on sample size and the difference in outcome between the groups being tested. In this study, the researchers want to detect a difference between drug B and standard of care if one exists; they want to maximize power.

β is the probability of committing a type II error **(Choices B and D)**. Type II error occurs when researchers fail to reject the null hypothesis when it is truly false. This causes investigators to miss true relationships. An example of a type II error would be a study finding that aspirin does not affect platelet function when, in fact, it does. Therefore, if β is set at 0.2, the power will be (1 – β) = 80%; there will be an 80% chance of rejecting the null hypothesis when it is truly false.

**(Choices A and C)** Type I error occurs when researchers reject the null hypothesis when the null hypothesis is really true. That is, a study finds a statistically significant difference between 2 groups when one does not truly exist. An example would be if a study concluded that candy improves heart failure mortality when, in fact, it does not. α is the maximum probability of making a type I error that a researcher is willing to accept. Generally, α is compared to the p-value, the probability of observing a given result (or more extreme) due to chance alone assuming the null hypothesis is true (eg, if there is no real difference between the groups). The value of α is typically set at 0.05, meaning that researchers are willing to accept up to a 5% chance of making a type I error. In such a scenario, if $p < 0.05$, the result is said to be statistically significant.

**Educational objective:**
Power (1 − β) is the probability of rejecting a null hypothesis when it is truly false. It is typically set at 80% and depends on sample size and difference between outcomes.

Time Spent: 6 seconds



Last updated: [8/17/2015]

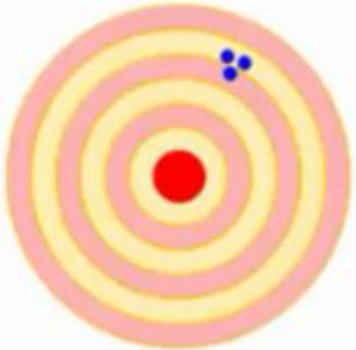
A new portable cholesterol measuring device determines a patient's blood cholesterol level to be 200 mg/dL on three separate measurements of the same blood sample. Using the gold standard measurement method, the same sample is found to have a cholesterol level of 260 mg/dL. The new measuring device is:

- A. Accurate but not precise
- B. Precise but not reliable
- C. Sensitive but not specific
- D. Reliable but not accurate
- E. Sensitive but not valid

A new portable cholesterol measuring device determines a patient's blood cholesterol level to be 200 mg/dL on three separate measurements of the same blood sample. Using the gold standard measurement method, the same sample is found to have a cholesterol level of 260 mg/dL. The new measuring device is:

- A. Accurate but not precise [11%]
- B. Precise but not reliable [22%]
- C. Sensitive but not specific [4%]
- ✔ D. Reliable but not accurate [60%]
- E. Sensitive but not valid [2%]

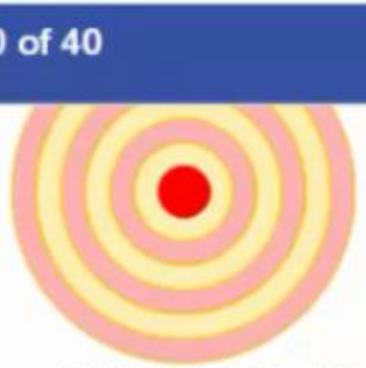
**Explanation:**



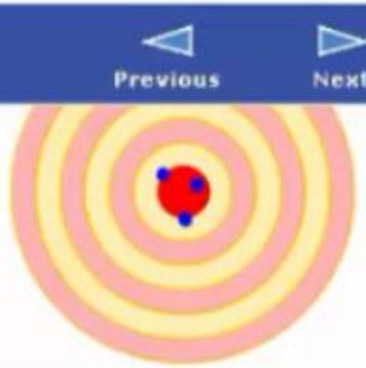
*Reliability (or) Precision*

*Accurate*

A *reliable* tool is one that consistently provides the same value when measuring a fixed entity. An *accurate* tool is one that gives a measurement close to the actual value (as reflected in the gold standard measurement)



*Reliability (or) Precision*



*Accurate*

A *reliable* tool is one that consistently provides the same value when measuring a fixed entity. An *accurate* tool is one that gives a measurement close to the actual value (as reflected in the gold standard measurement).

**(Choice A)** A tool with a high level of *precision* produces measurements with a low level of variability when used to assess a fixed entity. Precision and reliability are essentially equivalent terms. The measuring device described above is precise but not accurate.

**(Choice B)** The measuring device described above is both precise and reliable.

**(Choices C and E)** Sensitivity and specificity refer to the ability of a diagnostic test to discriminate between individuals who do and do not have a defined clinical condition. Validity is the extent to which a measurement actually measures what it claims to measure.

**Educational Objective:**
The reliability of a measurement technique refers to its reproducibility. The accuracy of a measurement technique is the degree to which the average measurement value matches that of the gold standard technique.

Time Spent: 9 seconds



Last updated: [8/19/2015]

A large multi-country study used population data from 14 countries located within similar latitudes to evaluate the association between dietary habits, including vitamin D intake, and prevalence of multiple sclerosis. Which of the following best describes the design of this study?

- A. Case-control study
- B. Cohort study
- C. Cross-sectional survey
- D. Ecological study
- E. Nested case-control study
- F. Qualitative study
- G. Randomized controlled trial
- H. Systematic review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

A large multi-country study used population data from 14 countries located within similar latitudes to evaluate the association between dietary habits, including vitamin D intake, and prevalence of multiple sclerosis. Which of the following best describes the design of this study?

- A. Case-control study [6%]
- B. Cohort study [13%]
- C. Cross-sectional survey [46%]
- ✔ D. Ecological study [26%]
- E. Nested case-control study [1%]
- F. Qualitative study [2%]
- G. Randomized controlled trial [1%]
- H. Systematic review [5%]

**Explanation:**

The unit of analysis in this study is populations not individuals. This is consistent with an **ecological study**, in which the frequency of a given characteristic (eg, vitamin D intake) and a given outcome (eg, multiple sclerosis [MS]) are studied using **population data**. Ecological studies are useful to generate hypotheses but should not be used to make conclusions regarding individuals within these populations (**ecological fallacy**).

**(Choices A and B)** Case-control and cohort studies would start with individuals rather than populations. In case-control studies, the odds of exposure to a certain

**(Choices A and B)** Case-control and cohort studies would start with individuals rather than populations. In case-control studies, the odds of exposure to a certain characteristic (eg, high or low vitamin D intake) is compared between affected individuals (eg, patients with MS) and unaffected individuals who serve as controls. In cohort studies, individuals (with and without different exposures such as high or low vitamin D intake) are followed over time to determine incidence of the disease of interest (such as MS).

**(Choice C)** Cross-sectional surveys would also evaluate the exposures and outcomes of interest in individuals (not populations) at a given point in time ("snapshot").

**(Choice E)** Nested case-control designs start with cohort studies in which participants are followed over time, and those participants who develop an outcome of interest become cases for a case-control study.

**(Choice F)** Qualitative studies use focused discussion groups, interviews (structured and semi-structured), and other anthropologic techniques to obtain narrative information that can be crucial in explaining quantitative results.

**(Choice G)** Randomized controlled trials enroll individuals who will be randomly assigned into a treatment group or a control group. The groups differ only in terms of the intervention (treatment) of interest.

**(Choice H)** Systematic reviews and meta-analyses take several studies (with an emphasis on high-quality randomized controlled studies) and attempt to group the results to obtain a pooled effect estimate.

**Educational objective:**
The unit of analysis in ecological studies is populations not individuals.

A study is designed to evaluate the efficacy of a new drug, KM28. The study will compare KM28 plus standard care versus standard care alone with regard to decreasing the incidence of recurrent breast cancer. The Federal Drug Administration (FDA) will approve the new drug if KM28 plus standard care decreases the rate of breast cancer recurrence by at least 40% compared to standard therapy alone. The recurrence rate on standard therapy is found to be 8%. In order for the FDA to approve KM28, what is the maximal incidence of recurrent disease acceptable for women treated with KM28 plus standard therapy?

- A. 2.8%
- B. 3.2%
- C. 3.6%
- D. 4.8%
- E. 5.2%

A study is designed to evaluate the efficacy of a new drug, KM28. The study will compare KM28 plus standard care versus standard care alone with regard to decreasing the incidence of recurrent breast cancer. The Federal Drug Administration (FDA) will approve the new drug if KM28 plus standard care decreases the rate of breast cancer recurrence by at least 40% compared to standard therapy alone. The recurrence rate on standard therapy is found to be 8%. In order for the FDA to approve KM28, what is the maximal incidence of recurrent disease acceptable for women treated with KM28 plus standard therapy?

- A. 2.8% [3%]
- B. 3.2% [24%]
- C. 3.6% [9%]
- ✔ D. 4.8% [60%]
- E. 5.2% [4%]

**Explanation:**

The new drug, KM28, will be approved if its associated recurrence rate is decreased by at least 40% compared to the recurrence rate on standard therapy alone, which is given as 8%. As 40% of 8% is 0.40 × 8% = 3.2%, the maximum acceptable recurrence rate is 8% − 3.2% = 4.8%. Another quick solution would be to state that the maximum acceptable recurrence rate is 60% of 8%, which is 0.60 × 8% = 4.8%.

An alternate solution involves using relative and absolute risk calculations. The question asks you to recognize that the 40% mentioned in the prompt refers to relative risk reduction (RRR). RRR is defined as the percent reduction in absolute

**Explanation:**

The new drug, KM28, will be approved if its associated recurrence rate is decreased by at least 40% compared to the recurrence rate on standard therapy alone, which is given as 8%. As 40% of 8% is 0.40 × 8% = 3.2%, the maximum acceptable recurrence rate is 8% − 3.2% = 4.8%. Another quick solution would be to state that the maximum acceptable recurrence rate is 60% of 8%, which is 0.60 × 8% = 4.8%.

An alternate solution involves using relative and absolute risk calculations. The question asks you to recognize that the 40% mentioned in the prompt refers to relative risk reduction (RRR). RRR is defined as the percent reduction in absolute risk (AR) between the treatment and control groups. The formula for RRR is:

$$RRR = (AR_{control} - AR_{treatment}) / AR_{control}$$

Values for $AR_{control}$ and RRR are provided in the question stem, so the next step is to solve for $AR_{treatment}$. The formula can thus be rearranged to:

$$AR_{treatment} = AR_{control} - (RRR \times AR_{control})$$

In this case, RRR is 0.4 and $AR_{control}$ is 0.08, which gives:

$$AR_{treatment} = 0.08 - (0.4 \times 0.08) = 0.08 - 0.032 = 0.048 \text{ (ie, 4.8\%)}$$

It is important to remember that RRR may overstate the effectiveness of an intervention. For example, a RRR of 50% occurs whether a drug decreases the incidence of a disease from 2% to 1% or from 50% to 25%. Clearly, the latter is of greater clinical significance.

**Educational objective:**
Relative risk reduction = (absolute $risk_{control}$ − absolute $risk_{treatment}$) / absolute $risk_{control}$

A rare metabolic disease has severe consequences if left untreated. If detected early after birth, the disease can be treated effectively and all the severe sequelae can be prevented. It is most important for a test to be used in all newborns to have a high:

- ○ A. Cutoff value
- ○ B. Number of true negatives
- ○ C. Positive predictive value
- ○ D. Sensitivity
- ○ E. Specificity

A rare metabolic disease has severe consequences if left untreated. If detected early after birth, the disease can be treated effectively and all the severe sequelae can be prevented. It is most important for a test to be used in all newborns to have a high:

- A. Cutoff value [1%]
- B. Number of true negatives [1%]
- C. Positive predictive value [9%]
- ✔ D. Sensitivity [79%]
- E. Specificity [10%]

**Explanation:**

The **sensitivity** of a test refers to its ability to correctly identify those with the disease; it is the probability of a positive test result in a person with the disease. Most patients with the disease will have a positive test result on a test with high sensitivity; this is important for **screening**. Given a test with high **sensitivity**, a negative result would help to **rule out** a diagnosis (**SnNout**). Such a test would be essential in this scenario for 2 reasons: 1) the metabolic disease has severe consequences if left untreated; 2) if detected early, the disease is treatable and the consequences are preventable.

**(Choice C)** The positive predictive value refers to the probability that a disease is present given a positive test result. Positive predictive value depends on disease prevalence.

**Explanation:**

The **sensitivity** of a test refers to its ability to correctly identify those with the disease; it is the probability of a positive test result in a person with the disease. Most patients with the disease will have a positive test result on a test with high sensitivity; this is important for **screening**. Given a test with high **sensitivity**, a negative result would help to **rule out** a diagnosis (**SnNout**). Such a test would be essential in this scenario for 2 reasons: 1) the metabolic disease has severe consequences if left untreated; 2) if detected early, the disease is treatable and the consequences are preventable.

**(Choice C)** The positive predictive value refers to the probability that a disease is present given a positive test result. Positive predictive value depends on disease prevalence.

**(Choices A, B, and E)** Specificity represents the ability of a test to correctly identify those without the disease. A very specific test has a low rate of false positives. This means that most healthy patients will have a negative test result (true negative). It is important for a confirmatory test to have a high specificity. Given a test with high specificity, a positive result would help to rule in a diagnosis (SpPin). Setting a high cutoff value typically produces higher specificity and lower sensitivity.

**Educational objective:**
The sensitivity of a test refers to its ability to correctly identify those with the disease. It is important to have high sensitivity in screening tests. Given a test with high sensitivity, a negative result would help to rule out a diagnosis (SnNout).

Time Spent: 6 seconds



Last updated: [8/2/2015]

A study is conducted assessing the relationship between smoking status and the forced expiratory flow between 25% and 75% of vital capacity ($FEF_{25-75\%}$) in 4 groups. Group A consists of 200 nonsmokers, group B consists of 200 light smokers (1-7 cigarettes per day), group C consists of 200 moderate smokers (8-22 cigarettes per day), and group D consists of 200 heavy smokers (23+ cigarettes per day). $FEF_{25-75\%}$ is measured in all participants. Which of the following is the most appropriate statistical method to compare the $FEF_{25-75\%}$ results among all 4 groups?

- A. Analysis of variance
- B. Chi-square test
- C. Meta-analysis
- D. Multiple linear regression
- E. Pearson correlation coefficient
- F. Two-sample t-test

A study is conducted assessing the relationship between smoking status and the forced expiratory flow between 25% and 75% of vital capacity ($FEF_{25-75\%}$) in 4 groups. Group A consists of 200 nonsmokers, group B consists of 200 light smokers (1-7 cigarettes per day), group C consists of 200 moderate smokers (8-22 cigarettes per day), and group D consists of 200 heavy smokers (23+ cigarettes per day). $FEF_{25-75\%}$ is measured in all participants. Which of the following is the most appropriate statistical method to compare the $FEF_{25-75\%}$ results among all 4 groups?

- ✔ (●) A. Analysis of variance [54%]
- ( ) B. Chi-square test [25%]
- ( ) C. Meta-analysis [8%]
- ( ) D. Multiple linear regression [6%]
- ( ) E. Pearson correlation coefficient [5%]
- ( ) F. Two-sample t-test [3%]

**Explanation:**

## Analysis of variance (ANOVA)

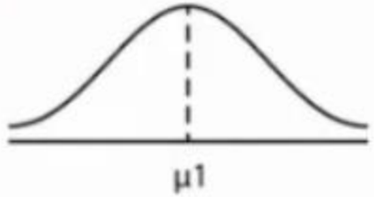
| Null hypothesis | Alternate hypothesis |
|---|---|
| All samples drawn from the same population ($\mu1 = \mu2 = \mu3$) | Samples drawn from different populations ($\mu1 \neq \mu2 \neq \mu3$) |

**Explanation:**

## Analysis of variance (ANOVA)

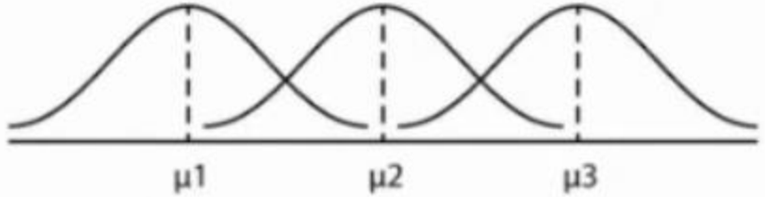
**Null hypothesis**

All samples drawn from the same population ($\mu1 = \mu2 = \mu3$)

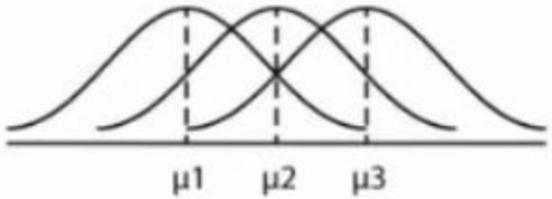



**Alternate hypothesis**

Samples drawn from different populations ($\mu1 \neq \mu2 \neq \mu3$)



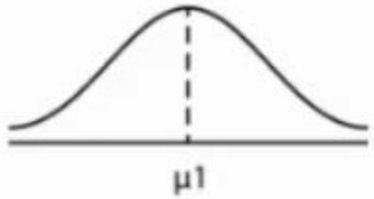

Small apparent difference between the means of the samples
null hypothesis not rejected




Large apparent differences between the means of the samples
null hypothesis rejected

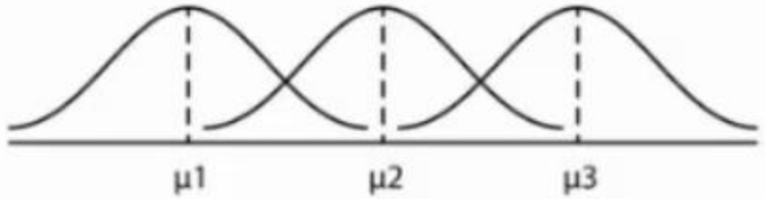
**Null hypothesis**

All samples drawn from the same population
($\mu 1 = \mu 2 = \mu 3$)

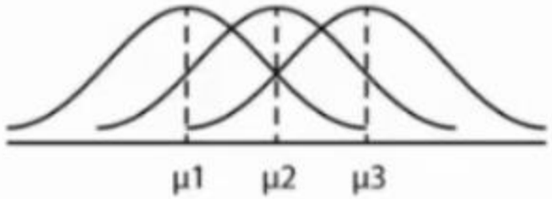



**Alternate hypothesis**

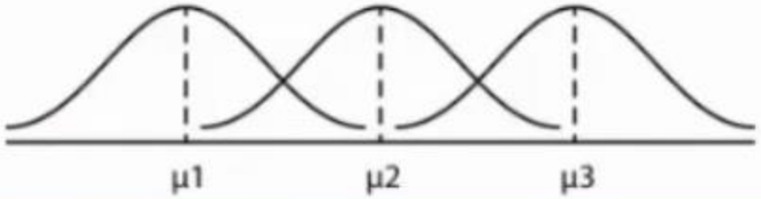
Samples drawn from different populations
($\mu 1 \neq \mu 2 \neq \mu 3$)




Small apparent difference between the means of the samples
null hypothesis not rejected




Large apparent differences between the means of the samples
null hypothesis rejected



Analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of 2 or more independent groups. ANOVA compares the means between the groups relative to the variability within groups and determines whether any of those means are significantly different from one another. Specifically, it tests the null hypothesis that all groups are simply random samples of the same population. The null hypothesis is rejected and the alternative hypothesis is accepted when there are at least 2 group means that are significantly different from one another.

**(Choice B)** Chi-square tests can be used to check for an association between 2 categorical variables. For example, a chi-square test could be used to determine if the distribution of gender and smoking status is random or if there is a difference between the sexes regarding smoking status.

**(Choice C)** Meta-analysis involves the pooling of data from several studies to perform an analysis having greater statistical power than the individual studies themselves. For example, individual studies assessing the effects of aspirin on certain cardiovascular events may be inconclusive. However, analysis of data compiled from multiple clinical trials may reveal a significant benefit.

**(Choice D)** Multiple linear regression is a method used to model the linear relationship between a dependent variable and 2 or more independent variables. For example, this test could be used to quantify the effects of alcohol use, tobacco smoking, and charred food consumption on the incidence of gastric cancer.

**(Choice E)** The Pearson correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between hemoglobin$_{A1c}$ levels and average blood glucose levels.

categorical variables. For example, a chi-square test could be used to determine if the distribution of gender and smoking status is random or if there is a difference between the sexes regarding smoking status.

**(Choice C)** Meta-analysis involves the pooling of data from several studies to perform an analysis having greater statistical power than the individual studies themselves. For example, individual studies assessing the effects of aspirin on certain cardiovascular events may be inconclusive. However, analysis of data compiled from multiple clinical trials may reveal a significant benefit.

**(Choice D)** Multiple linear regression is a method used to model the linear relationship between a dependent variable and 2 or more independent variables. For example, this test could be used to quantify the effects of alcohol use, tobacco smoking, and charred food consumption on the incidence of gastric cancer.

**(Choice E)** The Pearson correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between hemoglobin$_{A1c}$ levels and average blood glucose levels.

**(Choice F)** A two-sample t-test can be used when 2 group means need to be compared. This test could have been used for the example given in the question if the study participants were divided only into smoking and nonsmoking groups.

**Educational objective:**
A t-test is used to compare the difference between the means of 2 groups. Analysis of variance (ANOVA) compares the difference between the means of 2 or more groups.

Time Spent: 10 seconds



Last updated: [1/16/2016]

A state with a population of 4,000,000 contains 20,000 people who have disease A, a fatal neurodegenerative condition. There are 7000 new cases of the disease a year and 1000 deaths attributable to disease A. There are 40,000 deaths per year from all causes. What is the incidence of the disease?

- A. 1000 / 4,000,000
- B. 6000 / 4,000,000
- C. 7000 / 3,980,000
- D. 20,000 / 4,000,000
- E. 40,000 / 4,000,000

A state with a population of 4,000,000 contains 20,000 people who have disease A, a fatal neurodegenerative condition. There are 7000 new cases of the disease a year and 1000 deaths attributable to disease A. There are 40,000 deaths per year from all causes. What is the incidence of the disease?

- A. 1000 / 4,000,000 [1%]
- B. 6000 / 4,000,000 [6%]
- ✔ C. 7000 / 3,980,000 [84%]
- D. 20,000 / 4,000,000 [6%]
- E. 40,000 / 4,000,000 [1%]

**Explanation:**

The incidence of a disease is the number of new cases of a disease per year divided by the total population at risk. In this case it is 7000 / (4,000,000 - 20,000)

**(Choice A)** The disease-specific mortality is the number deaths attributable to a disease per year divided by the total population. It gives 1000/ 4,000,000.

**(Choice B)** The rate of increase of a disease is the number of new cases per year minus the number of death (or cures) per year, divided by the total population. This yields (7000-1000) / 4,000,000.

**(Choice D)** The prevalence of a disease is the number of persons with a disease divided by the total population at a specific point in time. In this case it is 20,000 / 4,000,000

1
2
3
4
• 5
• 6
• 7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

- B. 6000 / 4,000,000 [6%]
- ✔ C. 7000 / 3,980,000 [84%]
- D. 20,000 / 4,000,000 [6%]
- E. 40,000 / 4,000,000 [1%]

**Explanation:**

The incidence of a disease is the number of new cases of a disease per year divided by the total population at risk. In this case it is 7000 / (4,000,000 - 20,000)

**(Choice A)** The disease-specific mortality is the number deaths attributable to a disease per year divided by the total population. It gives 1000/ 4,000,000.

**(Choice B)** The rate of increase of a disease is the number of new cases per year minus the number of death (or cures) per year, divided by the total population. This yields (7000-1000) / 4,000,000.

**(Choice D)** The prevalence of a disease is the number of persons with a disease divided by the total population at a specific point in time. In this case it is 20,000 / 4,000,000

**(Choice E)** The mortality rate is the number of deaths per year divided by the total population. It yields 40,000 / 4,000,000.

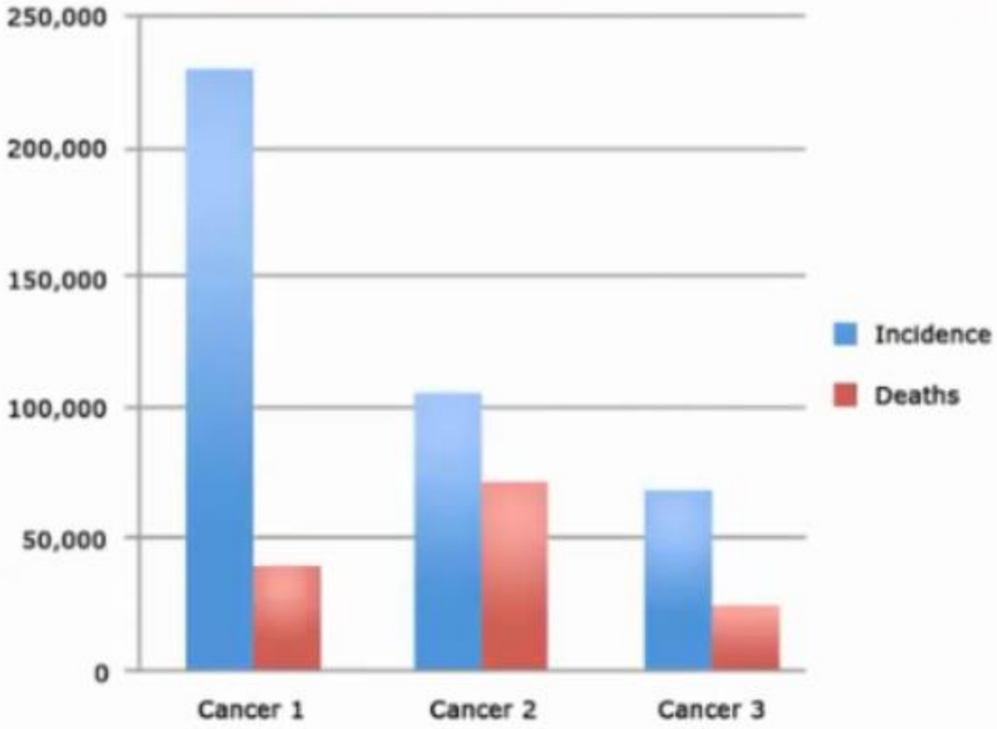
**Educational Objective:**
The incidence of a disease is the number of new cases of a disease per year divided by the total population at risk.

Time Spent: 10 seconds



Last updated: [8/19/2015]

The figure below shows the incidence and mortality for the three most common cancers affecting women living in the United States during 2011. The x-axis lists the specific cancers, and the y-axis shows the number of patients affected.

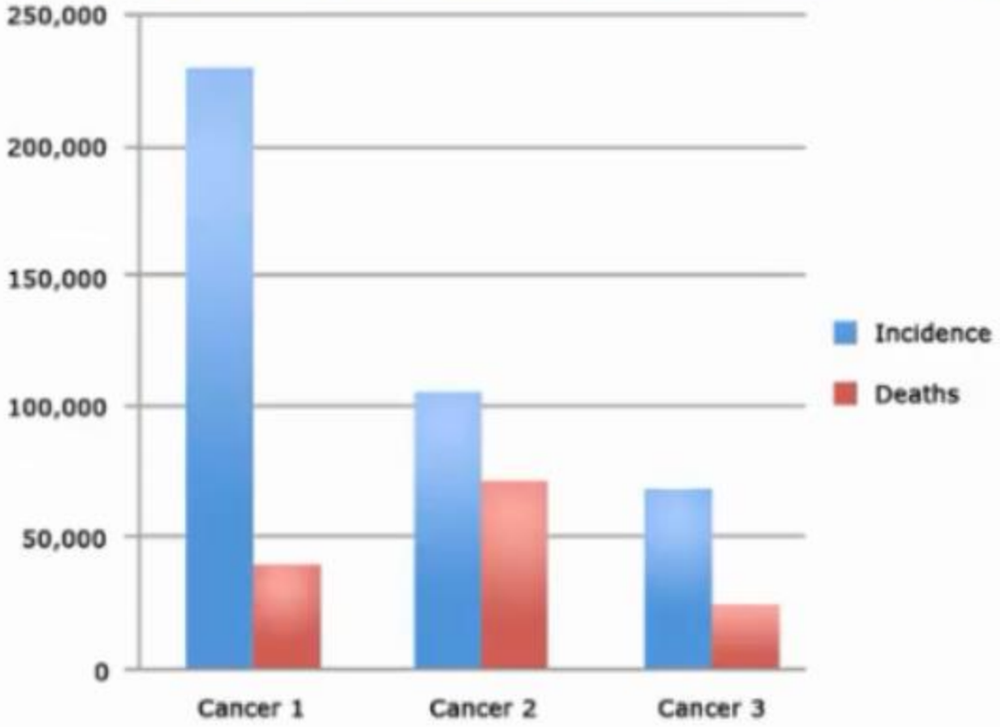



Which of the following lists of cancers best corresponds to cancers 1, 2, and 3 respectively?

- A. Breast, lung, colon
- B. Lung, breast, colon

1 2 3 4 5 • 6 • 7 • 8 • 9 • 10 • 11 • 12 • 13 • 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26

specific cancers, and the y-axis shows the number of patients affected.

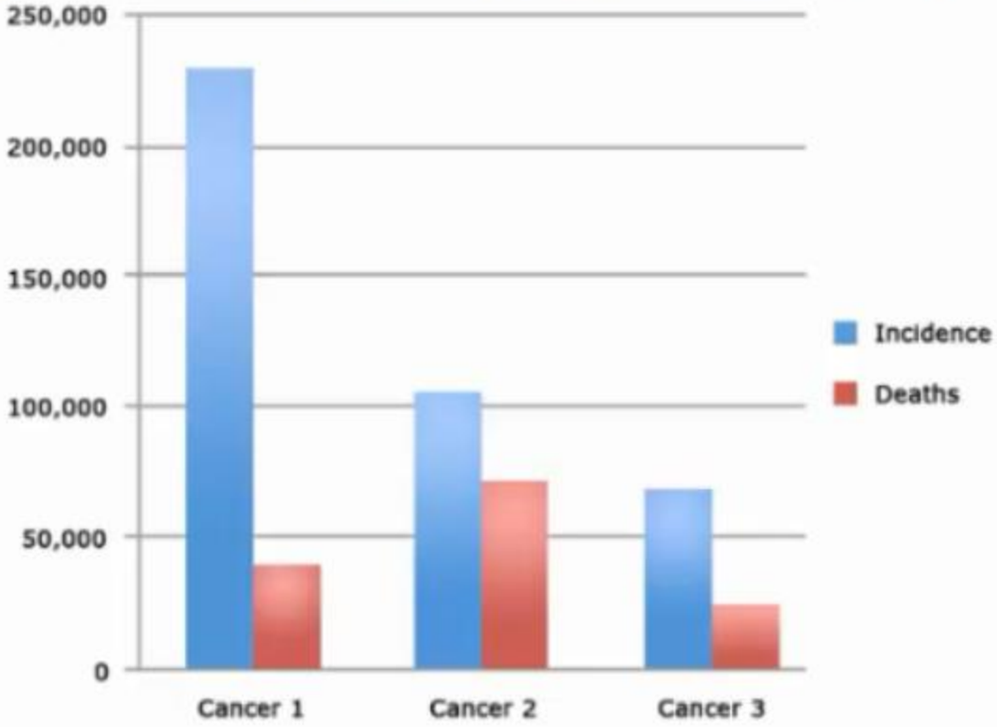



Which of the following lists of cancers best corresponds to cancers 1, 2, and 3 respectively?

- A. Breast, lung, colon
- B. Lung, breast, colon
- C. Breast, ovary, lung
- D. Lung, breast, cervix
- E. Breast, colon, lung

1
2
3
4
5
• 6
• 7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

specific cancers, and the y-axis shows the number of patients affected.



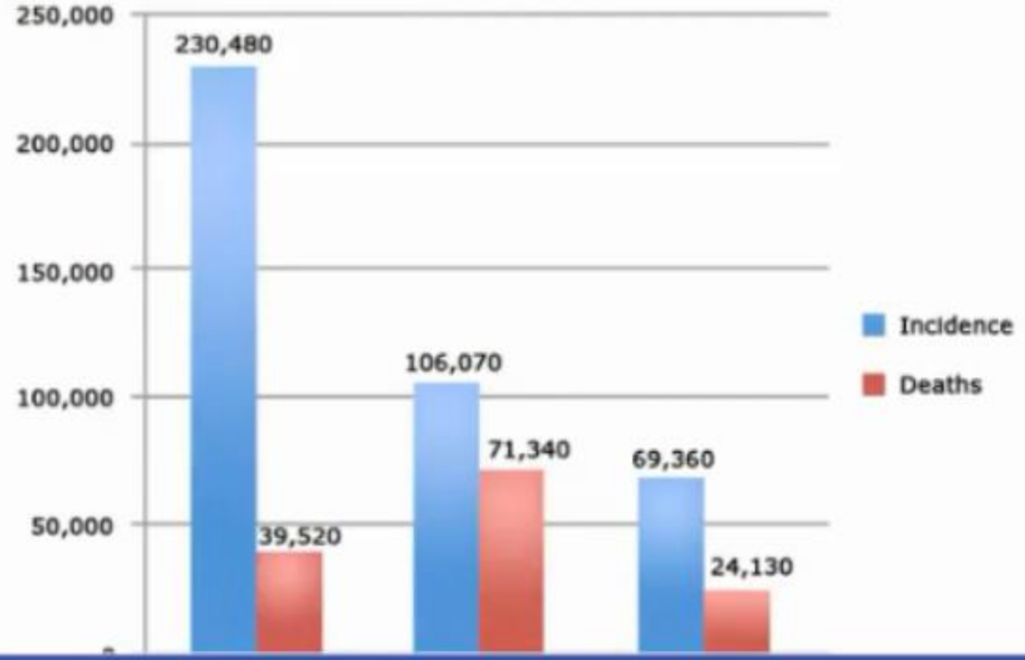

Which of the following lists of cancers best corresponds to cancers 1, 2, and 3 respectively?

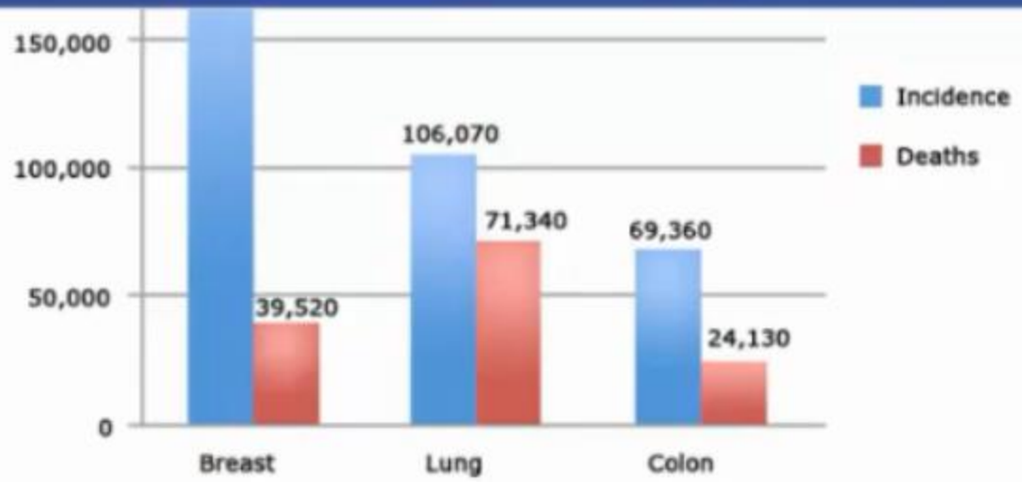
- ✔ (●) A. Breast, lung, colon [83%]
- ( ) B. Lung, breast, colon [8%]
- ( ) C. Breast, ovary, lung [3%]
- ( ) D. Lung, breast, cervix [3%]
- ( ) E. Breast, colon, lung [2%]

1
2
3
4
5
• 6
• 7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**

According to the American Cancer Society's *Cancer Facts & Figures 2011*, the most common cancers among women living in the United States during 2011 were breast cancer, lung cancer, and colon cancer. In order to arrange the cancer statistics in the correct order, a basic knowledge of cancer incidence and their corresponding mortality is needed. Aside from skin cancers, breast cancer is the most common cancer among women in the United States, but its mortality rate is moderately low. In fact, 230,480 breast cancers were diagnosed in the United States during 2011, but only 39,520 breast cancer deaths were reported. Despite the fact that the incidence of lung cancer is much lower than that for breast cancer, mortality from lung cancer is higher. Lung cancer has emerged as the leading cause of cancer mortality in women.






**(Choices C and D)** The incidences of ovarian and cervical cancers are much lower than that of the three cancers presented on the slide.

**(Choice E)** The incidence of colon cancer is somewhat close to that of lung, but the associated mortality is relatively much lower.

**Educational objective:**
According to 2011 statistics, the most common cancers (aside from skin cancer) in women in order of incidence were breast, lung, and colon cancer. In terms of mortality, lung cancer claimed the most lives, followed by breast and then colon cancer.

**References:**

1. Breast cancer statistics, 2011.

Time Spent: 26 seconds



Last updated: [1/15/2016]

Middle-aged men who have a first-degree relative with Alzheimer disease have their blood carotene concentration measured. They are followed over 20 years for the development of Alzheimer disease. The results are given below:

| | Low carotene level | Normal carotene level | Total |
|---|---|---|---|
| Develop Alzheimer disease | 20 | 40 | 60 |
| Do not develop Alzheimer disease | 20 | 120 | 140 |
| | 40 | 160 | 200 |

What is the 20-year risk of developing Alzheimer disease in subjects with low carotene level?

- A. 0.14
- B. 0.20
- C. 0.25
- D. 0.33
- E. 0.50

Middle-aged men who have a first-degree relative with Alzheimer disease have their blood carotene concentration measured. They are followed over 20 years for the development of Alzheimer disease. The results are given below:

| | Low carotene level | Normal carotene level | Total |
|---|---|---|---|
| Develop Alzheimer disease | 20 | 40 | 60 |
| Do not develop Alzheimer disease | 20 | 120 | 140 |
| | 40 | 160 | 200 |

What is the 20-year risk of developing Alzheimer disease in subjects with low carotene level?

- A. 0.14 [2%]
- B. 0.20 [8%]
- C. 0.25 [8%]
- D. 0.33 [14%]
- ✔ E. 0.50 [69%]

**Explanation:**

**Risk** is the probability of developing a disease or other health outcome over the study period. In this example, it represents the probability of developing Alzheimer disease over a 20-year period among middle-aged men who have a first-degree

1
2
3
4
5
6
• 7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**

**Risk** is the probability of developing a disease or other health outcome over the study period. In this example, it represents the probability of developing Alzheimer disease over a 20-year period among middle-aged men who have a first-degree relative with Alzheimer disease. To calculate this probability in subjects with low carotene levels, divide the number of subjects who develop Alzheimer disease in the low carotene group (20) by the overall number of subjects in the low carotene group (20 + 20 = 40).

> Risk of developing Alzheimer disease among subjects with low carotene = 20 / (20 + 20) = 20/40 = 0.50

This means that, among subjects with low carotene levels, there is a 50% probability of developing Alzheimer disease in 20 years.

Note that this is different from the **relative risk (RR)** of Alzheimer disease in subjects with low carotene levels compared to those with normal carotene groups:

> RR = Risk of developing Alzheimer disease among subjects with low carotene levels / Risk of developing Alzheimer disease among subjects with normal carotene levels = 0.5 / [40 / (40 + 120)] = 0.5/0.25 = 2

As the data is not presented in the standard 2 × 2 (contingency) table format, care must be taken to perform the calculations without relying on memorized formulas.

**(Choices A and D)** The prevalence of low carotene (exposure) among subjects who do not develop Alzheimer disease is 20 / (20 + 120) = 20/140 = 0.14 (14%). The prevalence of low carotene (exposure) among subjects who develop Alzheimer disease is 20 / (20 + 40) = 20 / 60 = 0.33 (33%).

This means that, among subjects with low carotene levels, there is a 50% probability of developing Alzheimer disease in 20 years.

Note that this is different from the **relative risk (RR)** of Alzheimer disease in subjects with low carotene levels compared to those with normal carotene groups:

RR = Risk of developing Alzheimer disease among subjects with low carotene levels / Risk of developing Alzheimer disease among subjects with normal carotene levels = 0.5 / [40 / (40 + 120)] = 0.5/0.25 = 2

As the data is not presented in the standard 2 × 2 (contingency) table format, care must be taken to perform the calculations without relying on memorized formulas.

**(Choices A and D)** The prevalence of low carotene (exposure) among subjects who do not develop Alzheimer disease is 20 / (20 + 120) = 20/140 = 0.14 (14%). The prevalence of low carotene (exposure) among subjects who develop Alzheimer disease is 20 / (20 + 40) = 20 / 60 = 0.33 (33%).

**(Choice B)** The prevalence of low carotene (exposure) in the entire cohort is (20 + 20) / (20 + 40 + 20 + 120) = 40/200 = 0.20 (20%).

**(Choice C)** As noted in the RR calculation above, the 20-year risk of developing Alzheimer disease among subjects with normal carotene levels is 40 / (40 + 120) = 40/160 = 0.25 (25%).

**Educational objective:**
Risk is the probability of developing a disease over a certain period of time. To calculate this probability, divide the number of affected subjects by the total number of subjects in the corresponding exposure group.

Time Spent: 6 seconds



Last updated: [1/18/2016]

1
2
3
4
5
6
7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A large cohort study is conducted to assess the association between smoking and squamous cell carcinoma of the esophagus among middle-aged Chinese men. During 10 years of follow-up, smokers have 5 times the risk of esophageal carcinoma compared to non-smokers (relative risk = 5.0, 95% confidence interval = 2.9-7.1). According to the study results, what percentage of squamous cell carcinoma of the esophagus in smokers can be attributed to smoking?

- ○ A. 25%
- ○ B. 50%
- ○ C. 70%
- ○ D. 80%
- ○ E. 90%

1
2
3
4
5
6
7
• 8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A large cohort study is conducted to assess the association between smoking and squamous cell carcinoma of the esophagus among middle-aged Chinese men. During 10 years of follow-up, smokers have 5 times the risk of esophageal carcinoma compared to non-smokers (relative risk = 5.0, 95% confidence interval = 2.9-7.1). According to the study results, what percentage of squamous cell carcinoma of the esophagus in smokers can be attributed to smoking?

- A. 25% [13%]
- B. 50% [13%]
- C. 70% [11%]
- ✔ D. 80% [52%]
- E. 90% [11%]

**Explanation:**

The **attributable risk percent in the exposed ($ARP_{exposed}$)** is an important measure of the impact of a risk factor. $ARP_{exposed}$ represents the excess risk in an exposed population that can be explained by exposure to a particular risk factor. It is calculated using the following formula:

**$ARP_{exposed}$ = 100 × [(risk in exposed – risk in unexposed)/risk in exposed]**

This basic definition can be used to derive an equivalent formula involving relative risk (RR):

**$ARP_{exposed}$ = 100 × [(RR – 1)/RR]**, where RR = risk in exposed/risk in

**Explanation:**

The **attributable risk percent in the exposed ($ARP_{exposed}$)** is an important measure of the impact of a risk factor. $ARP_{exposed}$ represents the excess risk in an exposed population that can be explained by exposure to a particular risk factor. It is calculated using the following formula:

**$ARP_{exposed}$ = 100 × [(risk in exposed – risk in unexposed)/risk in exposed]**

This basic definition can be used to derive an equivalent formula involving relative risk (RR):

**$ARP_{exposed}$ = 100 × [(RR – 1)/RR]**, where RR = risk in exposed/risk in unexposed

Applying the formula to this example:

$ARP_{exposed}$ = 100 × [(RR – 1)/RR] = 100 × [(5 – 1)/5] = 100 × (4/5) = 100 × 0.8 = 80%

Therefore, according to the study results, 80% of esophageal squamous cell carcinoma cases in smokers were attributable to smoking.

The $ARP_{exposed}$ is related to the attributable risk (AR), which is simply the difference between risk in the exposed and risk in the unexposed.

**Educational objective:**
The attributable risk percent (ARP) in the exposed represents the excess risk in the exposed population that can be attributed to the risk factor. It can be easily derived from the relative risk (RR) using the formula: $ARP_{exposed}$ = 100 × [(RR – 1)/RR].

A large prospective study evaluates the relationship between alcohol consumption and breast cancer. A total of 4000 middle-aged women are enrolled in the study through a random selection of residential addresses. Daily alcohol consumption and breast cancer incidence are assessed through the use of periodic questionnaires. Five-year follow-up shows that alcohol consumers are somewhat more likely to develop breast cancer (relative risk = 1.32, 95% confidence interval = 0.90-1.85). The investigators also report that 800 subjects were lost to follow-up by the end of the study, the majority of whom were moderate to heavy alcohol consumers. According to this information, which of the following biases is most likely to be present and may have affected the results?

- A. Lead-time bias
- B. Observer bias
- C. Random misclassification bias
- D. Recall bias
- E. Selection bias

1
2
3
4
5
6
7
8
• 9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A large prospective study evaluates the relationship between alcohol consumption and breast cancer. A total of 4000 middle-aged women are enrolled in the study through a random selection of residential addresses. Daily alcohol consumption and breast cancer incidence are assessed through the use of periodic questionnaires. Five-year follow-up shows that alcohol consumers are somewhat more likely to develop breast cancer (relative risk = 1.32, 95% confidence interval = 0.90-1.85). The investigators also report that 800 subjects were lost to follow-up by the end of the study, the majority of whom were moderate to heavy alcohol consumers. According to this information, which of the following biases is most likely to be present and may have affected the results?

- ○ A. Lead-time bias [15%]
- ○ B. Observer bias [7%]
- ○ C. Random misclassification bias [11%]
- ○ D. Recall bias [11%]
- ✔ ◉ E. Selection bias [56%]

**Explanation:**

**Attrition bias**

## Explanation:

**Bias** refers to a systematic error (due to nonrandom factors) that can result in an incorrect association between the exposure and outcome. In general, when bias is present, the results are consistently distorted in one direction. In prospective studies, if loss to follow-up occurs disproportionately between the exposed and unexposed groups, **attrition bias** can result if the lost subjects differ in their risk of developing the outcome compared to the remaining subjects. Attrition bias is a form of **selection bias** (a term that most often refers to systematic differences between groups in terms of treatment response or prognosis). Attrition bias does not occur when the losses happen randomly between the exposed and unexposed groups, as this simply leads to a smaller study population.

In this example, a substantial number of subjects (800 / 4000 = 20%) were lost to follow-up, primarily those with higher levels of alcohol consumption. If these lost subjects had a higher incidence of breast cancer than those remaining in the study, then a selective loss of high-risk subjects would have occurred in the exposed group. As a result, the measure of the association between alcohol use and breast cancer would be underestimated, which might explain why no significant association was seen in this study.

**(Choice A)** Lead-time bias occurs when a screening test diagnoses a disease earlier than it would have appeared by natural history alone, so that the time from diagnosis until death appears prolonged even though there might actually be no improvement in survival.

**(Choice B)** Observer bias can occur in clinical trials when study participants or investigators are aware of individual treatment assignments. It can be prevented by performing a double-blind study in which neither participants nor investigators are aware of treatment assignments.

diagnosis until death appears prolonged even though there might actually be no improvement in survival.

**(Choice B)** Observer bias can occur in clinical trials when study participants or investigators are aware of individual treatment assignments. It can be prevented by performing a double-blind study in which neither participants nor investigators are aware of treatment assignments.

**(Choice C)** Misclassification bias occurs when either the exposure or the outcome is not identified correctly. Random (or nondifferential) misclassification affects all groups to the same extent. For example, if a pediatric size sphygmomanometer cuff is used on all participants (treatment and control groups) as part of a study in adult patients, the blood pressure readings will be incorrect (due to the incorrect cuff size). However, the resulting misclassification will likely affect both groups to the same extent (because the same cuff was used).

**(Choice D)** Recall bias results from the inaccurate recall of past exposure status. It is a potential problem for retrospective studies such as case-control studies, particularly when questionnaires are used to inquire about distant past exposure. However, in this prospective study, exposure status is determined through periodic questionnaires assessing daily alcohol consumption (without inquiring about distant past exposure). In such a scenario, recall bias is less likely to affect results.

**Educational objective:**
In prospective studies, disproportionate loss to follow-up between the exposed and unexposed groups creates the potential for attrition bias, which is a form of selection bias. As a result, investigators generally try to achieve high patient follow-up rates in prospective studies.

Time Spent: 8 seconds



Last updated: [8/19/2015]

1
2
3
4
5
6
7
8
9
• 10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A new serologic test for detecting prostate cancer is negative in 95% of patients who do not have the disease. If the test is used on 8 blood samples taken from patients without prostate cancer, what is the probability of getting at least 1 positive test result?

- A. 1 − (0.05 * 8)
- B. 0.05 * 8
- C. 0.05^8
- D. 0.95^8
- E. (1 − 0.95^8)

A new serologic test for detecting prostate cancer is negative in 95% of patients who do not have the disease. If the test is used on 8 blood samples taken from patients without prostate cancer, what is the probability of getting at least 1 positive test result?

- A. 1 – (0.05 * 8) [21%]
- B. 0.05 * 8 [35%]
- C. 0.05^8 [9%]
- D. 0.95^8 [3%]
- ✔ E. (1 – 0.95^8) [31%]

**Explanation:**

Each of the 8 blood sample results is an **independent event** (one patient's results has no impact on another's) with a 0.95 (95%) probability of (correctly) testing negative and a 0.05 (5%) probability of (incorrectly) testing positive with the new serologic test. Overall, there are 2 possible outcomes: **either** all 8 samples will test negative **or** at least 1 sample will test positive (ie, 1 sample could test positive, or 2, or 3, or so on). These outcomes represent **complementary** events because their probabilities add up to 1 (100%) (exhaustive) and they cannot occur simultaneously (mutually exclusive). Mathematically:

P(all 8 samples test negative) + P(at least 1 sample tests positive) = 1, or equivalently,
P(**at least 1** sample tests positive) = 1 – P(all 8 samples test negative)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

**Explanation:**

Each of the 8 blood sample results is an **independent event** (one patient's results has no impact on another's) with a 0.95 (95%) probability of (correctly) testing negative and a 0.05 (5%) probability of (incorrectly) testing positive with the new serologic test. Overall, there are 2 possible outcomes: **either** all 8 samples will test negative **or** at least 1 sample will test positive (ie, 1 sample could test positive, or 2, or 3, or so on). These outcomes represent **complementary** events because their probabilities add up to 1 (100%) (exhaustive) and they cannot occur simultaneously (mutually exclusive). Mathematically:

> P(all 8 samples test negative) + P(at least 1 sample tests positive) = 1, or equivalently,
> P(**at least 1** sample tests positive) = 1 – P(all 8 samples test negative)

Using the **multiplication rule** for independent events:
P(all 8 samples test negative) = P(1st sample tests negative) * P(2nd sample tests negative) * … * P(7th sample tests negative) * P(8th sample tests negative) = 0.95 * 0.95 * … * 0.95 * 0.95 = 0.95^8.

Therefore, P(at least 1 sample tests positive) = 1 – P(all 8 samples test negative) = (1 – 0.95^8).

**(Choices A and B)** The value 0.05 * 8 was obtained by adding the probability of a sample testing positive 8 times (0.05 + 0.05 + … + 0.05 = 0.05 * 8). However, probabilities for a series of independent events should be obtained by multiplication (0.05^8), not addition. If 1 ball is pulled from 8 bowls, each with 95 white and 5 black balls: P(8 white balls pulled) = 0.95 * 0.95 * … * 0.95 * 0.95 = 0.95^8 = 0.66 (66%) by the multiplication rule. Addition is incorrect as 0.95 + 0.95 + … + 0.95 + 0.95 = 0.95 * 8 = 7.6; this is a 760% probability that exceeds 100%.

P(**at least 1** sample tests positive) = 1 – P(all 8 samples test negative)

Using the **multiplication rule** for independent events:
P(all 8 samples test negative) = P(1st sample tests negative) * P(2nd sample tests negative) * … * P(7th sample tests negative) * P(8th sample tests negative) = 0.95 * 0.95 * … * 0.95 * 0.95 = 0.95^8.

Therefore, P(at least 1 sample tests positive) = 1 – P(all 8 samples test negative) = (1 – 0.95^8).

**(Choices A and B)** The value 0.05 * 8 was obtained by adding the probability of a sample testing positive 8 times (0.05 + 0.05 + … + 0.05 = 0.05 * 8). However, probabilities for a series of independent events should be obtained by multiplication (0.05^8), not addition. If 1 ball is pulled from 8 bowls, each with 95 white and 5 black balls: P(8 white balls pulled) = 0.95 * 0.95 * … * 0.95 * 0.95 = 0.95^8 = 0.66 (66%) by the multiplication rule. Addition is incorrect as 0.95 + 0.95 + … + 0.95 + 0.95 = 0.95 * 8 = 7.6; this is a 760% probability that exceeds 100%.

**(Choices C and D)** P(all 8 samples test positive) = P(1st sample tests positive) * P (2nd sample tests positive) * … * P(8th sample tests positive) = 0.05^8. Note that 0.05^8 is mathematically equivalent to (1 – 0.95)^8 but not to (1 – 0.95^8), based on the order of operations indicated by parentheses. As calculated above, 0.95^8 is the probability that all samples test negative.

**Educational objective:**
If events are independent, the probability that all events will turn out the same is the product of the separate probabilities for each event. The probability of at least 1 event turning out differently is given as 1 – P(all events being the same).

Time Spent: 11 seconds



Last updated: [12/14/2015]

1
2
3
4
5
6
7
8
9
10
• 11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A prospective study evaluates the relationship between regular antioxidant use (vitamin C and vitamin E) and risk of stroke in men age 40-60. The results of the study were adjusted to account for confounding variables. According to the study results, the men who consumed antioxidants for at least 5 years have a stroke relative risk of 0.75 ($p < 0.01$). In contrast, the men who consumed antioxidants for less than 5 years have a stroke relative risk of 0.95 ($p = 0.45$). Which of the following factors most likely explains why the relative risk of stroke is lower with longer antioxidant use?

- A. Latent period
- B. Lead-time bias
- C. Observer bias
- D. Rare disease assumption
- E. Selection bias

A prospective study evaluates the relationship between regular antioxidant use (vitamin C and vitamin E) and risk of stroke in men age 40-60. The results of the study were adjusted to account for confounding variables. According to the study results, the men who consumed antioxidants for at least 5 years have a stroke relative risk of 0.75 ($p < 0.01$). In contrast, the men who consumed antioxidants for less than 5 years have a stroke relative risk of 0.95 ($p = 0.45$). Which of the following factors most likely explains why the relative risk of stroke is lower with longer antioxidant use?

- ✔ (●) A. Latent period [32%]
- ( ) B. Lead-time bias [35%]
- ( ) C. Observer bias [8%]
- ( ) D. Rare disease assumption [4%]
- ( ) E. Selection bias [21%]

**Explanation:**

This study reveals that a minimum of 5 years of antioxidant use is associated with a 0.75 relative risk (RR) of stroke, a finding that is statistically significant ($p < 0.01$). The use of antioxidants for <5 years demonstrated a smaller decrease in risk (RR of 0.95). However, the p-value of 0.45 for this result indicates that this small reduction in risk is not statistically significant. Therefore, the use of antioxidants for <5 years seems to have no appreciable benefit on stroke risk reduction. These results are explained by assuming that stroke risk reduction through antioxidant use is associated with a long latent period.

**Explanation:**

This study reveals that a minimum of 5 years of antioxidant use is associated with a 0.75 relative risk (RR) of stroke, a finding that is statistically significant ($p < 0.01$). The use of antioxidants for <5 years demonstrated a smaller decrease in risk (RR of 0.95). However, the p-value of 0.45 for this result indicates that this small reduction in risk is not statistically significant. Therefore, the use of antioxidants for <5 years seems to have no appreciable benefit on stroke risk reduction. These results are explained by assuming that stroke risk reduction through antioxidant use is associated with a long latent period.

Latency is a very important issue to consider when studying disease epidemiology. In most infectious diseases, the **latent period** (ie, time elapsed from initial exposure to clinically apparent disease) is relatively short. On the other hand, some disease processes (eg, cancer or heart disease) may demonstrate a very long latent period before clinical manifestations develop. The concept of a latency period can also be extended to risk factors and risk reducers. Exposure to risk modifiers sometimes occurs a significant amount of time before the exposure's effect on the disease process is clinically evident. Additionally, sometimes exposure to risk modifiers must occur continuously over extended periods of time before the disease outcome is affected. In this case, at least 5 years of continuous exposure to antioxidants were required to reveal the protective effect of the exposure on the outcome (stroke).

**(Choice B)** Lead time is the time between the initial detection of a disease and a specific outcome or measured endpoint. Lead-time bias can occur when a test diagnoses or detects the disease at an earlier stage than another test does without impacting the natural history of the disease. A study comparing disease survival times may then erroneously conclude that using the earlier-detection test prolongs survival, when in actuality, the increased survival time is due solely to earlier

**(Choice B)** Lead time is the time between the initial detection of a disease and a specific outcome or measured endpoint. Lead-time bias can occur when a test diagnoses or detects the disease at an earlier stage than another test does without impacting the natural history of the disease. A study comparing disease survival times may then erroneously conclude that using the earlier-detection test prolongs survival, when in actuality, the increased survival time is due solely to earlier detection of the disease.

**(Choice C)** Observer bias occurs when an observer misclassifies data due to individual differences in interpretation or preconceived expectations regarding a study. It can be reduced by performing a double blind study (both observers and participants unaware of randomized parameters) and by having multiple observers encode and verify the recorded data.

**(Choice D)** The rare disease assumption refers to the fact that the odds ratio approximates RR when disease prevalence is low.

**(Choice E)** Selection bias can occur with inappropriate (non-random) selection methods or through selective attrition of the study participants. Although selection bias is a possible limitation of the study (individuals who take vitamins daily may be more likely to be healthy), it does not explain the difference in risk of stroke over time as latency does.

**Educational objective:**
The concept of a latent period can be applied to both disease pathogenesis and exposure to risk modifiers. The initial steps in pathogenesis and/or exposure to a risk factor sometimes occur years before clinical manifestations of a disease are evident. Additionally, exposure to risk modifiers may need to be continuous over a certain period of time before influencing the outcome.

1
2
3
4
5
6
7
8
9
10
11
• 12
• 13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

A prospective cohort study was conducted to assess the role of daily alcohol consumption in the occurrence of breast carcinoma. The investigators reported a 5-year relative risk of 1.4 for people who consume alcohol daily compared to those who do not. The 95% confidence interval was 1.02-1.85. Which of the following p-values is most consistent with the results described above?

- A. 0.03
- B. 0.06
- C. 0.09
- D. 0.11
- E. 0.20

A prospective cohort study was conducted to assess the role of daily alcohol consumption in the occurrence of breast carcinoma. The investigators reported a 5-year relative risk of 1.4 for people who consume alcohol daily compared to those who do not. The 95% confidence interval was 1.02-1.85. Which of the following p-values is most consistent with the results described above?

- ✔ (●) A. 0.03 [69%]
- ( ) B. 0.06 [15%]
- ( ) C. 0.09 [7%]
- ( ) D. 0.11 [4%]
- ( ) E. 0.20 [6%]

**Explanation:**

Relative risk (RR) is used in cohort studies to determine how strongly a risk factor (ie, exposure) is associated with an outcome. RR is the risk of an outcome (eg, breast cancer) in the exposed group (eg, individuals who consume alcohol daily) divided by the risk of that outcome in the unexposed group (eg, individuals who do not consume alcohol daily). If the RR = 1.0 (null value), then there is no association between the exposure and the disease. An RR >1.0 indicates that the exposure is associated with increased risk of disease. An RR <1.0 means that the exposure is associated with decreased risk of disease.

The RR by itself does not account for the possibility that chance alone is responsible for the results. The 95% confidence interval (CI) and p-value are 2 measures of **statistical significance** that can help strengthen the findings of a study using RR.

**Explanation:**

Relative risk (RR) is used in cohort studies to determine how strongly a risk factor (ie, exposure) is associated with an outcome. RR is the risk of an outcome (eg, breast cancer) in the exposed group (eg, individuals who consume alcohol daily) divided by the risk of that outcome in the unexposed group (eg, individuals who do not consume alcohol daily). If the RR = 1.0 (null value), then there is no association between the exposure and the disease. An RR >1.0 indicates that the exposure is associated with increased risk of disease. An RR <1.0 means that the exposure is associated with decreased risk of disease.

The RR by itself does not account for the possibility that chance alone is responsible for the results. The 95% confidence interval (CI) and p-value are 2 measures of **statistical significance** that can help strengthen the findings of a study using RR. For a result to be considered statistically significant, its corresponding CI must NOT contain the null value. When the **95% CI** does **not** include the null value, this gives a corresponding **p-value <0.05** and the association between exposure and outcome is considered statistically significant. A p-value <0.05 reflects that there is a very low probability that the result was due to chance alone; formally, the p-value is the probability of observing a given (or more extreme) result due to chance alone assuming that the null hypothesis is true.

In this example, the RR is 1.4 with a 95% CI of 1.02-1.85. It can be concluded that daily alcohol consumption is associated with an increased risk of breast carcinoma (RR >1) and that the findings are statistically significant (95% CI does not include the null value of 1.0). Therefore, the expected p-value would be <0.05.

**(Choices B, C, D, and E)** These options contain p-values >0.05, so the results would not be statistically significant. Note the relationship between CI and

associated with increased risk of disease. An RR <1.0 means that the exposure is associated with decreased risk of disease.

The RR by itself does not account for the possibility that chance alone is responsible for the results. The 95% confidence interval (CI) and p-value are 2 measures of **statistical significance** that can help strengthen the findings of a study using RR. For a result to be considered statistically significant, its corresponding CI must NOT contain the null value. When the **95% CI** does **not** include the null value, this gives a corresponding **p-value <0.05** and the association between exposure and outcome is considered statistically significant. A p-value <0.05 reflects that there is a very low probability that the result was due to chance alone; formally, the p-value is the probability of observing a given (or more extreme) result due to chance alone assuming that the null hypothesis is true.

In this example, the RR is 1.4 with a 95% CI of 1.02-1.85. It can be concluded that daily alcohol consumption is associated with an increased risk of breast carcinoma (RR >1) and that the findings are statistically significant (95% CI does not include the null value of 1.0). Therefore, the expected p-value would be <0.05.

**(Choices B, C, D, and E)** These options contain p-values >0.05, so the results would not be statistically significant. Note the relationship between CI and p-value: a statistically significant 95% CI corresponds to a p-value <0.05; a statistically significant 99% CI (would also not include the null value and likely be narrower than a 95% CI) corresponds to a p-value <0.01.

**Educational objective:**
A result is considered statistically significant if the 95% confidence interval does not cross the null value, which corresponds to a p-value <0.05.

Time Spent: 6 seconds



Last updated: [7/29/2015]

A researcher is interested in assessing the blood folate level of women who live in a population with a high incidence of neural tube defects. She takes a large random sample (n) of women age 18-45 and measures their blood folate levels. She reports the mean, standard deviation (SD), and variance. She concludes that she is 95% confident that the population mean of the blood folate level lies somewhere between 2.4 and 4.6 ng/mL. This last statement is based on which of the following calculations?

- ○ A. Mean ± 1.96*SD
- ○ B. Mean ± 1.96*SD/√n
- ○ C. Mean ± 2.58*SD
- ○ D. Mean ± 2.58*SD/√n
- ○ E. Mean ± SD/n

A researcher is interested in assessing the blood folate level of women who live in a population with a high incidence of neural tube defects. She takes a large random sample (n) of women age 18-45 and measures their blood folate levels. She reports the mean, standard deviation (SD), and variance. She concludes that she is 95% confident that the population mean of the blood folate level lies somewhere between 2.4 and 4.6 ng/mL. This last statement is based on which of the following calculations?

- A. Mean ± 1.96*SD [22%]
- ✔ B. Mean ± 1.96*SD/√n [58%]
- C. Mean ± 2.58*SD [7%]
- D. Mean ± 2.58*SD/√n [8%]
- E. Mean ± SD/n [5%]

**Explanation:**

The normal distribution of individual observations of a variable (eg, blood folate levels) can be described using the mean and **standard deviation (SD)** of these observations. For example, the mean +/- 1.96*SD would cover 95% of the observations and the mean +/- 2.58*SD would cover 99% of the observations (with 1.96 and 2.58 representing z-scores for 95% and 99% of the distribution, respectively). However, most research is done on **samples** rather than on the entire population. This introduces some variability (when different samples are drawn from the same population, the results could be slightly different).

The **standard error (SE)** is defined as SD/√n and is a way to account for this

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

**Explanation:**

The normal distribution of individual observations of a variable (eg, blood folate levels) can be described using the mean and **standard deviation (SD)** of these observations. For example, the mean +/- 1.96*SD would cover 95% of the observations and the mean +/- 2.58*SD would cover 99% of the observations (with 1.96 and 2.58 representing z-scores for 95% and 99% of the distribution, respectively). However, most research is done on **samples** rather than on the entire population. This introduces some variability (when different samples are drawn from the same population, the results could be slightly different).

The **standard error (SE)** is defined as SD/√n and is a way to account for this variability due to sampling. The corresponding confidence interval (CI) represents the range of values within which one can be confident that the true mean of the underlying population falls. In this case, a 95% CI would be given by the mean +/- 1.96*SE = mean +/- 1.96*SD/√n. For example, the range between 2.4 and 4.6 ng/mL represents the 95% CI (given by mean +/- 1.96*SD/√n) and allows the investigators to be 95% confident that the mean folate level in the underlying population lies somewhere in that range. The 99% CI would be given by the mean +/- 2.58*SE = mean +/- 2.58*SD/√n **(Choice D)**. Note that the sample size n is a component of the calculation of SE. As n increases, SE decreases, so the CI will become narrower and more precise.

**(Choices A and C)** As noted above, the SD reflects the spread of individual values in a normal distribution. The mean +/- 1.96*SD would cover 95% of the observations and the mean +/- 2.58*SD would cover 99% of the observations.

In journals, means are often reported as either mean +/- 1 SD or mean +/- SE. The sample size should be reported so that both values can always be deduced.

1.96 and 2.58 representing z-scores for 95% and 99% of the distribution, respectively). However, most research is done on **samples** rather than on the entire population. This introduces some variability (when different samples are drawn from the same population, the results could be slightly different).

The **standard error (SE)** is defined as SD/√n and is a way to account for this variability due to sampling. The corresponding confidence interval (CI) represents the range of values within which one can be confident that the true mean of the underlying population falls. In this case, a 95% CI would be given by the mean +/- 1.96*SE = mean +/- 1.96*SD/√n. For example, the range between 2.4 and 4.6 ng/mL represents the 95% CI (given by mean +/- 1.96*SD/√n) and allows the investigators to be 95% confident that the mean folate level in the underlying population lies somewhere in that range. The 99% CI would be given by the mean +/- 2.58*SE = mean +/- 2.58*SD/√n **(Choice D)**. Note that the sample size n is a component of the calculation of SE. As n increases, SE decreases, so the CI will become narrower and more precise.

**(Choices A and C)** As noted above, the SD reflects the spread of individual values in a normal distribution. The mean +/- 1.96*SD would cover 95% of the observations and the mean +/- 2.58*SD would cover 99% of the observations.

In journals, means are often reported as either mean +/- 1 SD or mean +/- SE. The sample size should be reported so that both values can always be deduced.

**Educational objective:**
The standard deviation reflects the spread of individual values in a normal distribution. The standard error of the mean reflects the variability of means and helps estimate the true mean of the underlying population.

Time Spent: 11 seconds



Last updated: [1/18/2016]

1
2
3
4
5
6
7
8
9
10
11
12
13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

Kidney biopsy samples from 500 diabetic patients were examined by several pathologists. The pathologists that knew the diabetes status of patients were three times more likely to interpret the sample microscopy findings as "diabetic nephropathy." Which of the following most likely explains the difference in interpreting the microscopy results?

- A. Confounding
- B. Lead-time bias
- C. Recall bias
- D. Selection bias
- E. Observer bias

Kidney biopsy samples from 500 diabetic patients were examined by several pathologists. The pathologists that knew the diabetes status of patients were three times more likely to interpret the sample microscopy findings as "diabetic nephropathy." Which of the following most likely explains the difference in interpreting the microscopy results?

- A. Confounding [4%]
- B. Lead-time bias [2%]
- C. Recall bias [6%]
- D. Selection bias [6%]
- ✔ E. Observer bias [82%]

**Explanation:**

Observer bias occurs when the investigator's decision is affected by prior knowledge of the exposure status. In this case, some pathologists were prejudiced by the clinical history because they knew that diabetic nephropathy is a common finding in patients with diabetes. The pathologists who were blinded to the patients' diabetes status were not prejudiced by the clinical history and had unbiased interpretations.

**(Choice A)** Confounding is present when at least part of the exposure-disease relationship can be explained by another variable (confounder). No information on possible confounders is given in this scenario.

**(Choice B)** Lead-time bias refers to the apparent prolongation of survival in patients

1
2
3
4
5
6
7
8
9
10
11
12
13
• 14
• 15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

✔ E. Observer bias [82%]

**Explanation:**

Observer bias occurs when the investigator's decision is affected by prior knowledge of the exposure status. In this case, some pathologists were prejudiced by the clinical history because they knew that diabetic nephropathy is a common finding in patients with diabetes. The pathologists who were blinded to the patients' diabetes status were not prejudiced by the clinical history and had unbiased interpretations.

**(Choice A)** Confounding is present when at least part of the exposure-disease relationship can be explained by another variable (confounder). No information on possible confounders is given in this scenario.

**(Choice B)** Lead-time bias refers to the apparent prolongation of survival in patients who underwent a screening test that allowed for earlier diagnosis but did not improve prognosis.

**(Choice C)** Recall bias results from inaccurate patient recall of past exposure. It is not applicable in this case.

**(Choice D)** Selection bias can result when subjects are selected for a study or from selective losses during follow-up.

**Educational Objective:**
Observer bias occurs when the investigator's decision is affected by prior knowledge of the exposure status.

Time Spent: 10 seconds



Last updated: [8/19/2015]

A new biomarker has been shown to allow for the early detection of non-small cell lung carcinoma. A study of this new test demonstrates that its use prolongs survival of lung cancer patients by three months when compared to the survival of those subjects diagnosed by conventional methods. The researchers conclude that the use of this new biomarker improves prognosis in patients with non-small cell lung cancer. Which of the following is a potential problem with this conclusion?

- A. Observer bias
- B. Measurement bias
- C. Rare disease assumption
- D. Confounding
- E. Lead-time bias

A new biomarker has been shown to allow for the early detection of non-small cell lung carcinoma. A study of this new test demonstrates that its use prolongs survival of lung cancer patients by three months when compared to the survival of those subjects diagnosed by conventional methods. The researchers conclude that the use of this new biomarker improves prognosis in patients with non-small cell lung cancer. Which of the following is a potential problem with this conclusion?

- A. Observer bias [6%]
- B. Measurement bias [6%]
- C. Rare disease assumption [3%]
- D. Confounding [8%]
- ✔ E. Lead-time bias [77%]

**Explanation:**

The prospect of lead-time bias should always be considered when evaluating any screening test. Lead-time bias is defined as an artificial increase in survival time among tested patients who actually have an unchanged prognosis. Patients screened with more sensitive tests appear to live longer only because the disease was detected earlier than it would have been if diagnosed clinically. The overall length of time from disease onset to death actually remains the same in both groups.

**(Choices A and B)** Measurement bias and observer bias refer to the misclassification of outcome and/or exposure (eg, labeling diseased as non-diseased and vice versa) and are related to poor study design. The scenario described does not mention how the study was designed.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 • 15 • 16 • 17 • 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26

✔ ◉ E. Lead-time bias [77%]

**Explanation:**

The prospect of lead-time bias should always be considered when evaluating any screening test. Lead-time bias is defined as an artificial increase in survival time among tested patients who actually have an unchanged prognosis. Patients screened with more sensitive tests appear to live longer only because the disease was detected earlier than it would have been if diagnosed clinically. The overall length of time from disease onset to death actually remains the same in both groups.

**(Choices A and B)** Measurement bias and observer bias refer to the misclassification of outcome and/or exposure (eg, labeling diseased as non-diseased and vice versa) and are related to poor study design. The scenario described does not mention how the study was designed.

**(Choice D)** Although the results of the study could be potentially confounded, there is no information on how potential confounders were treated during the design or analysis stage of the study.

**(Choice C)** Rare disease assumption refers to the practice of approximating the odds ratio and relative risk when studying rare diseases.

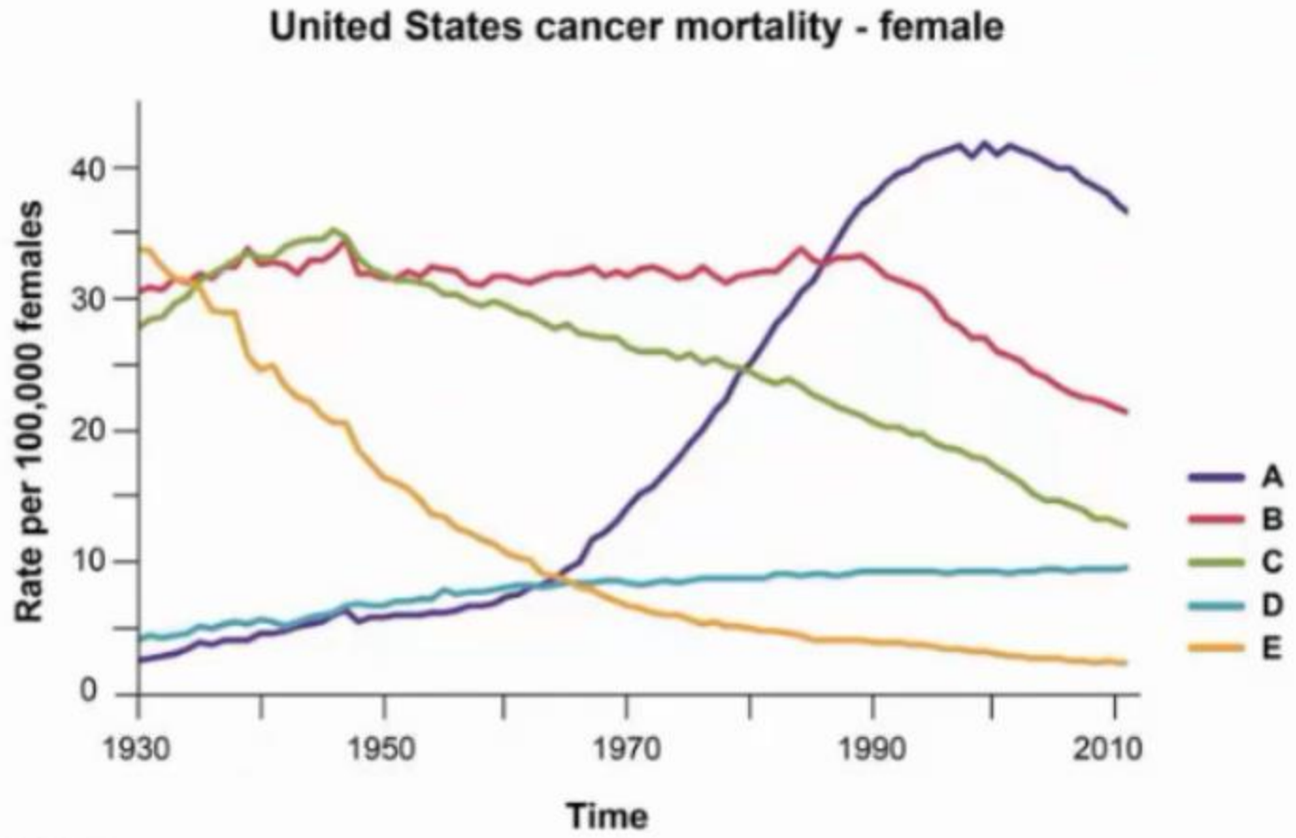
**Educational Objective:**
The typical example of lead-time bias is an apparent increase in survival in patients diagnosed with a new test who actually have an unchanged prognosis. Think of lead-time bias when you see "a new screening test" for poor prognosis diseases like lung or pancreatic cancer!

Time Spent: 7 seconds



Last updated: [8/19/2015]

The age-adjusted mortality trends for 5 cancers in women in the United States are shown in the graph below.

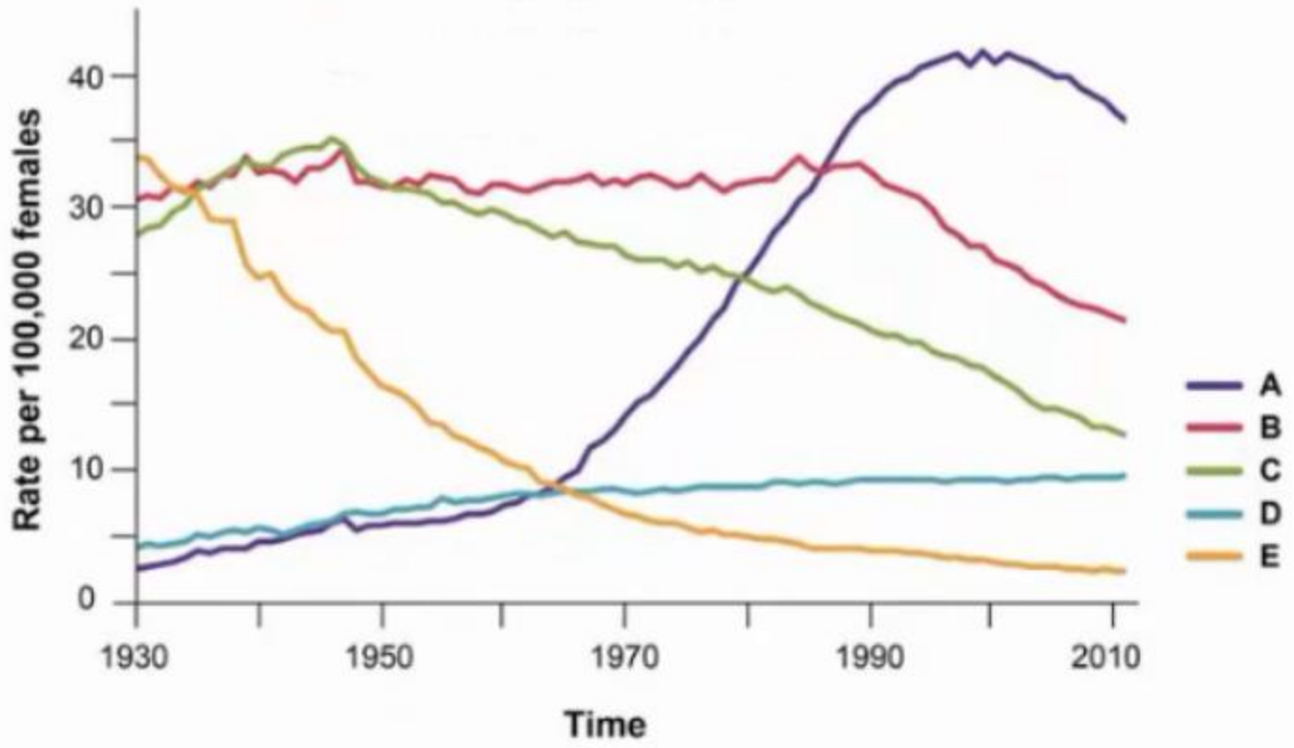





Which of the following curves best corresponds to lung cancer?

A. A
B. B

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

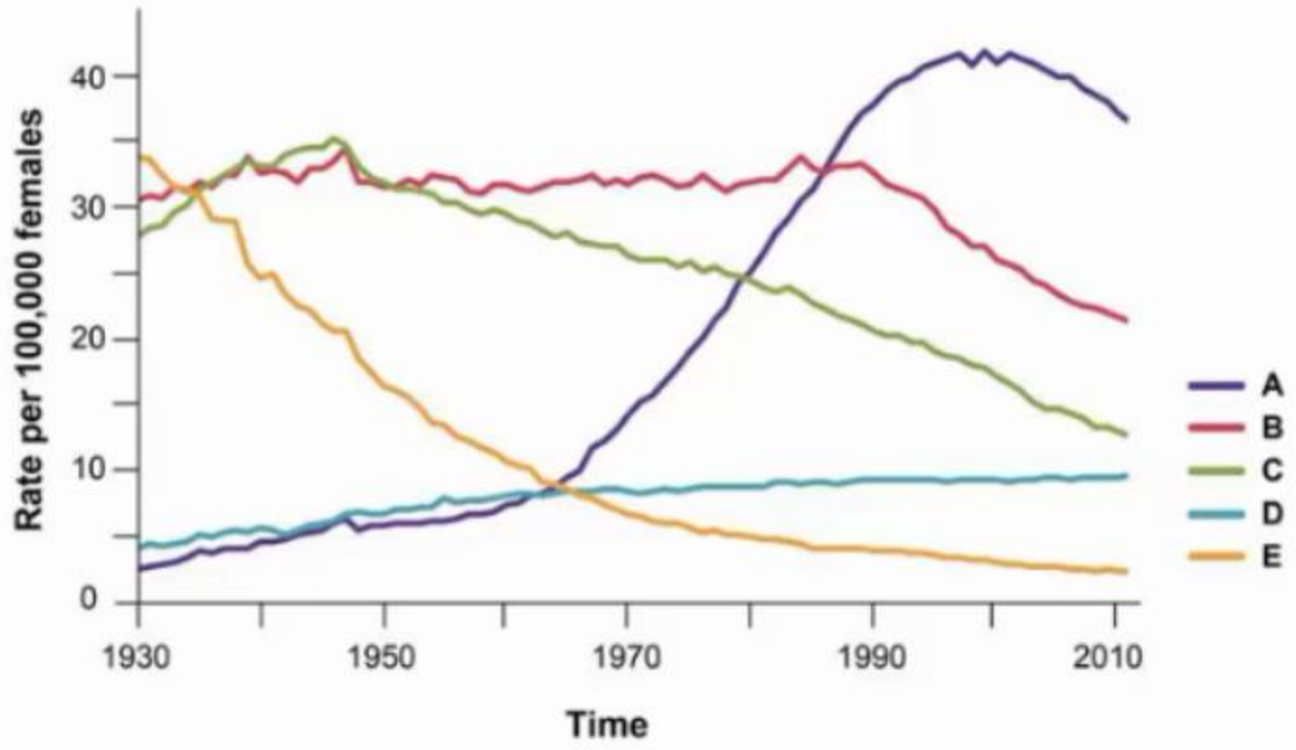





Which of the following curves best corresponds to lung cancer?

- A. A
- B. B
- C. C
- D. D
- E. E

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26



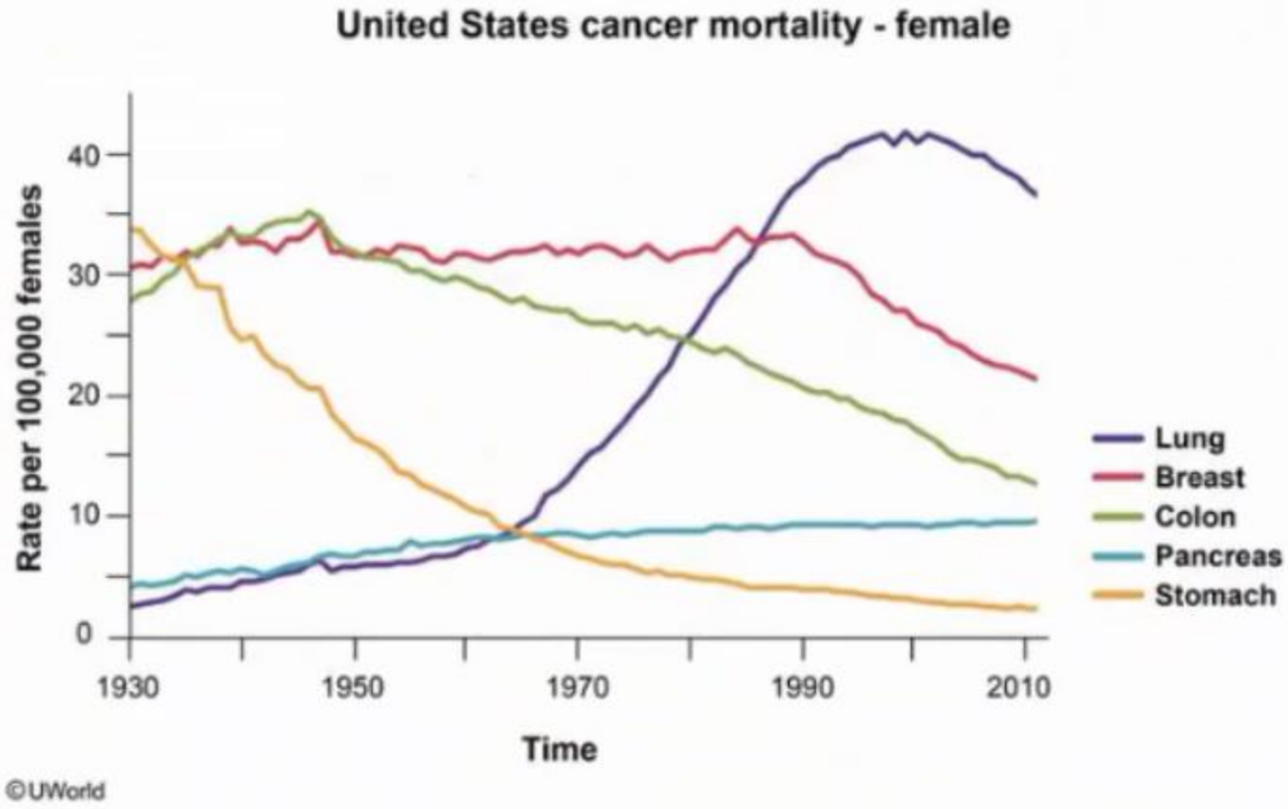



Which of the following curves best corresponds to lung cancer?

- ✔ A. A [76%]
- B. B [13%]
- C. C [3%]
- D. D [7%]
- E. E [1%]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
• 16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**




**Lung cancer** has been the leading cause of **cancer mortality** in both **women** and men in the United States since the 1980s. **Tobacco use** (primary and secondhand) is the most important cause of lung cancer, particularly **non-small cell** lung cancer. Female use of cigarettes peaked in 1955, and mortality rates increased 20-50 years after smoking onset. Lung cancer mortality began to decrease after 2000, corresponding to a decline in tobacco use.

Tobacco is a causative factor for numerous other cancers (eg, leukemia, nasopharyngeal, larynx, esophagus, pancreas, cervix, colon). It is directly carcinogenic to tissue by causing irritation and inflammation of the body's natural protective barriers. Patient education about tobacco's harmful effects is extremely important, as smoking cessation for even longtime users can be beneficial for reducing cancer risk and improving life expectancy.

**(Choice B)** Breast cancer is the most common non-skin cancer and the second most common cause of cancer death among women in the United States. Mortality from breast cancer began to decrease in the 1990s. Increased use of adjuvant chemotherapy and/or radiation and more frequent breast cancer screening are likely contributing factors to this decline.

**(Choice C)** Colon cancer mortality has decreased since the 1950s due to advances in surgical technique and adjuvant chemotherapy. Additional protective factors may include colorectal cancer screening, menopausal hormone therapy in women, and aspirin use.

**(Choice D)** The incidence and mortality of pancreatic cancer have increased in women over the last century and recently became the fourth most common cause of cancer death in women. Pancreatic cancer trends, like those of lung cancer, follow the pattern of increased smoking in women. However, lung cancer is responsible for three times more deaths than pancreatic cancer in women.

**(Choice E)** Stomach cancer incidence and mortality decreased drastically over the first half of the twentieth century. Proposed reasons include advances in refrigeration and food preservation (leading to decreased salt intake), and better sanitation and more adequate housing (reducing *Helicobacter pylori* infection rates).

**Educational objective:**

include colorectal cancer screening, menopausal hormone therapy in women, and aspirin use.

**(Choice D)** The incidence and mortality of pancreatic cancer have increased in women over the last century and recently became the fourth most common cause of cancer death in women. Pancreatic cancer trends, like those of lung cancer, follow the pattern of increased smoking in women. However, lung cancer is responsible for three times more deaths than pancreatic cancer in women.

**(Choice E)** Stomach cancer incidence and mortality decreased drastically over the first half of the twentieth century. Proposed reasons include advances in refrigeration and food preservation (leading to decreased salt intake), and better sanitation and more adequate housing (reducing *Helicobacter pylori* infection rates).

**Educational objective:**
Between 1950 and 2000, rising rates of tobacco use resulted in an increase in female lung cancer incidence and mortality. Lung cancer is currently the most common cause of cancer death in both women and men in the United States.

**References:**

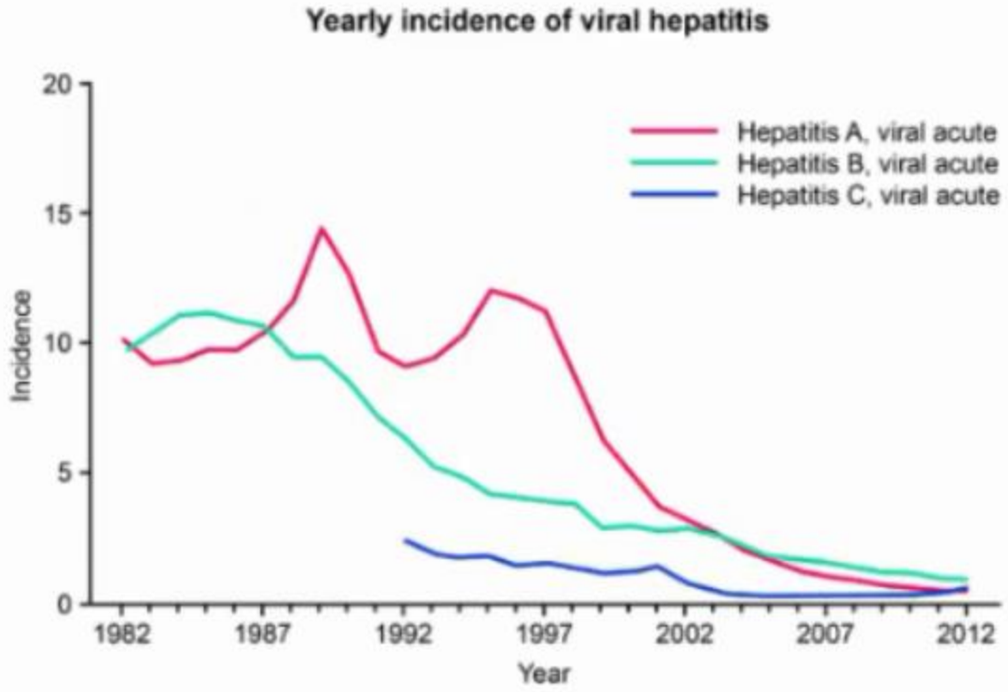
1. Lung cancer incidence trends among men and women--United States, 2005-2009.
2. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000.
3. Long-term trends in cancer mortality in the United States, 1930-1998.
4. Deaths: Final data for 2011.

Time Spent: 23 seconds



Last updated: [1/5/2016]

The graph below shows the yearly incidence of viral hepatitis per 100,000 population from 1982-2012 in the United States (US).

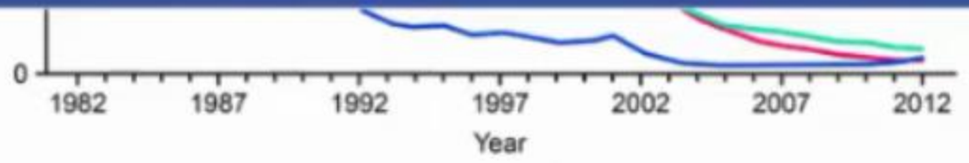



Viral hepatitis incidence (per 100,000 population) in United States, 1982-2012



During this period, the total population in the US continued to increase, as shown in the table below:

| Year | 1982 | 1987 | 1992 | 1997 | 2002 | 2007 | 2012 |
|---|---|---|---|---|---|---|---|
| US population (in millions) | ~230 | ~243 | ~258 | ~275 | ~290 | ~303 | ~315 |




Viral hepatitis incidence (per 100,000 population) in United States, 1982-2012

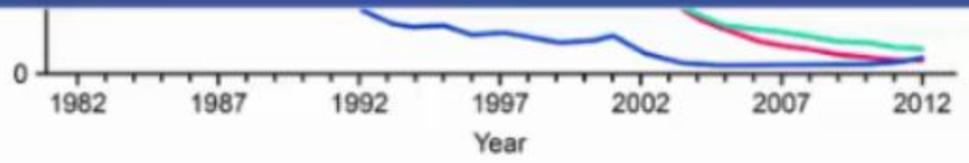


During this period, the total population in the US continued to increase, as shown in the table below:

| Year | 1982 | 1987 | 1992 | 1997 | 2002 | 2007 | 2012 |
|---|---|---|---|---|---|---|---|
| US population (in millions) | ~230 | ~243 | ~258 | ~275 | ~290 | ~303 | ~315 |

At any given time point, the number of individuals at risk for hepatitis infection is assumed to be equal to the total US population. Based on these data, which of the following is most likely to be true?

- A. After 2006, hepatitis B prevalence surpassed hepatitis A and hepatitis C prevalence
- B. In 1987, there were as many individuals with hepatitis A as with hepatitis B in the US
- C. In 1989, there were more individuals with hepatitis A than with hepatitis B in the US
- D. In 1997, there were more new cases of hepatitis A than of hepatitis B and C combined
- E. The number of new cases of hepatitis A diagnosed in 1982 and 1998 were exactly the same




Viral hepatitis incidence (per 100,000 population) in United States, 1982-2012



During this period, the total population in the US continued to increase, as shown in the table below:

| Year | 1982 | 1987 | 1992 | 1997 | 2002 | 2007 | 2012 |
|---|---|---|---|---|---|---|---|
| US population (in millions) | ~230 | ~243 | ~258 | ~275 | ~290 | ~303 | ~315 |

At any given time point, the number of individuals at risk for hepatitis infection is assumed to be equal to the total US population. Based on these data, which of the following is most likely to be true?

- A. After 2006, hepatitis B prevalence surpassed hepatitis A and hepatitis C prevalence [7%]
- B. In 1987, there were as many individuals with hepatitis A as with hepatitis B in the US [6%]
- C. In 1989, there were more individuals with hepatitis A than with hepatitis B in the US [10%]
- ✔ D. In 1997, there were more new cases of hepatitis A than of hepatitis B and C combined [72%]
- E. The number of new cases of hepatitis A diagnosed in 1982 and 1998 were exactly the same [5%]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
• 17
• 18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**

**Incidence** corresponds to the number of **new** cases of a disease diagnosed in a population at risk over a given time period. In 1997, incidence was >10 new cases/100,000 population for hepatitis A virus (HAV), <5/100,000 for hepatitis B virus (HBV), and <3/100,000 for hepatitis C virus (HCV). Assuming, as instructed, that the entire United States (US) population is at risk (reasonable theory given low number of hepatitis cases among the total population), then, in 1997 (US population ~275 million):

New HAV cases in 1997 = >10 new cases/100,000 population × 275,000,000 population = >27,500
New HBV cases in 1997 = <5/100,000 × 275,000,000 = <13,750
New HCV cases in 1997 = <3/100,000 × 275,000,000 = <8,250

There were thousands more new HAV (>27,500) than HBV and HCV cases combined (<13,750 + <8,250 = <22,000). The calculations incorporating total population (included for completion) are not needed to answer the question.

After 2006, HBV incidence surpassed HAV and HCV incidence. However, no information can be deduced regarding **prevalence**, which is the **total** number of cases in the population **(Choice A)**. Compared to new HBV cases, there were approximately just as many **new** HAV cases in 1987 and more new HAV cases in 1989. However, the **total** number of HAV or HBV cases in the population cannot be calculated without additional information. In fact, there were likely more total cases (ie, higher prevalence) of HBV as it can progress to a chronic form **(Choices B and C)**.

A simplified diagram compares incidence and prevalence to adding new drops (incident cases) into a water-containing sink (prevalent cases).

calculated without additional information. In fact, there were likely more total cases (ie, higher prevalence) of HBV as it can progress to a chronic form **(Choices B and C)**.

A simplified diagram compares incidence and prevalence to adding new drops (incident cases) into a water-containing sink (prevalent cases).

**(Choice E)** HAV incidence **per 100,000 population** was similar in 1982 and 1998. However, the overall US population, and consequently the number of new HAV cases, was different. Assuming HAV incidence was ~10 new cases/100,000 population for both years, then, in 1982 (US population ~230 million):

New HAV cases in 1982 = 10/100,000 × 230,000,000 = 23,000

Also, in 1998 (US population ~275 million at least):

New HAV cases in 1998 = 10/100,000 × 275,000,000 = 27,500

Even with different incidences, it is unlikely that exactly the same number of new HAV diagnoses were made.

**Educational objective:**
Incidence corresponds to the number of new cases of a disease in a certain population at risk over a given time period. Prevalence is the total number of cases in the population over a given period.

**References:**

1. Summary of notifiable diseases--United States, 2012.

Time Spent: 20 seconds



Last updated: [1/21/2016]

Women who have just delivered babies with neural tube defects are asked about their use of acetaminophen during the first three months of pregnancy. Women who delivered apparently healthy babies are also asked about their use of acetaminophen during the first three months of pregnancy. Which of the following measures of association are the investigators most likely to report?

- A. Relative risk
- B. Median survival
- C. Exposure odds ratio
- D. Relative rate
- E. Prevalence odds ratio

Women who have just delivered babies with neural tube defects are asked about their use of acetaminophen during the first three months of pregnancy. Women who delivered apparently healthy babies are also asked about their use of acetaminophen during the first three months of pregnancy. Which of the following measures of association are the investigators most likely to report?

- A. Relative risk [34%]
- B. Median survival [1%]
- ✔ C. Exposure odds ratio [58%]
- D. Relative rate [3%]
- E. Prevalence odds ratio [4%]

**Explanation:**

This scenario describes a typical case-control study design. People with the disease of interest (cases) and people without this disease (controls) are asked about previous exposure to the variable being studied (acetaminophen use). The main measure of association is the exposure odds ratio, which is expressed as the following equation:

$$\frac{\text{(Odds of exposure of people with the disease [cases])}}{\text{(Odds of exposure of people without the disease [controls])}}$$

**(Choices A and D)** Incidence measures (eg, relative risk or relative rate) cannot be directly measured in case-control studies because the people being studied are

**Explanation:**

This scenario describes a typical case-control study design. People with the disease of interest (cases) and people without this disease (controls) are asked about previous exposure to the variable being studied (acetaminophen use). The main measure of association is the exposure odds ratio, which is expressed as the following equation:

$$\frac{\text{(Odds of exposure of people with the disease [cases])}}{\text{(Odds of exposure of people without the disease [controls])}}$$

**(Choices A and D)** Incidence measures (eg, relative risk or relative rate) cannot be directly measured in case-control studies because the people being studied are those who already have the disease. Relative risk and relative rate are calculated in cohort studies, in which people are followed over time for developing the disease.

**(Choice B)** Median survival is calculated in cohort studies or clinical trials, and is usually used to compare the median survival times in two or more groups of patients (eg, those receiving a new treatment and those receiving a placebo).

**(Choice E)** Prevalence odds ratio is calculated in cross-sectional studies to compare the prevalence of a disease in different populations.

**Educational Objective:**
A case-control study is used to compare the exposure of people with the disease (cases) to the exposure of people without the disease (controls). The main measure of association is the exposure odds ratio.

Time Spent: 8 seconds



Last updated: [8/19/2015]

A geneticist is interested in the potential causes of a congenital abnormality. She selects the mothers of children with and without the abnormality randomly from a national birth registry. When interviewing the mothers of children with the abnormality, she discovers that several of them used acetaminophen during pregnancy. Mothers of children who do not have the abnormality report that they did not take acetaminophen as frequently. This type of investigation is most susceptible to which of the following types of bias?

- A. Referral bias
- B. Detection bias
- C. Lead-time bias
- D. Allocation bias
- E. Recall bias
- F. Sampling bias

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 • 19 • 20 • 21 • 22 • 23 • 24 • 25 • 26

A geneticist is interested in the potential causes of a congenital abnormality. She selects the mothers of children with and without the abnormality randomly from a national birth registry. When interviewing the mothers of children with the abnormality, she discovers that several of them used acetaminophen during pregnancy. Mothers of children who do not have the abnormality report that they did not take acetaminophen as frequently. This type of investigation is most susceptible to which of the following types of bias?

- A. Referral bias [1%]
- B. Detection bias [5%]
- C. Lead-time bias [2%]
- D. Allocation bias [2%]
- ✔ E. Recall bias [82%]
- F. Sampling bias [9%]

**Explanation:**

Recall bias results from inaccurate recall of past exposure by people in the study and applies mostly to retrospective studies such as case-control studies. People who have suffered an adverse event (such as having a child with congenital abnormalities) are more likely to recall previous risk factors than are people who have not experienced a poor outcome. The risk is known at the time of enrollment in prospective studies, so recall bias is eliminated.

**(Choice A)** Referral (admission rate) bias occurs when the case and control populations differ due to admission or referral practices. For instance, a study

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
• 19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**

Recall bias results from inaccurate recall of past exposure by people in the study and applies mostly to retrospective studies such as case-control studies. People who have suffered an adverse event (such as having a child with congenital abnormalities) are more likely to recall previous risk factors than are people who have not experienced a poor outcome. The risk is known at the time of enrollment in prospective studies, so recall bias is eliminated.

**(Choice A)** Referral (admission rate) bias occurs when the case and control populations differ due to admission or referral practices. For instance, a study involving cancer risk factors performed at a hospital specializing in cancer research may enroll cases referred from all over the nation. However, hospitalized control subjects without cancer may come from only the local area.

**(Choice B)** Detection bias refers to the fact that a risk factor itself may lead to extensive diagnostic investigation and increase the probability that a disease is identified. For instance, patients who smoke may undergo increased imaging surveillance due to their smoking status, which would detect more cases of cancer in general.

**(Choice C)** Lead-time bias occurs when a screening test diagnoses a disease earlier than it would have appeared by natural history so that the time from diagnosis until death is prolonged. Affected studies may not accurately reflect an improvement in survival or alteration of the natural history.

**(Choice D)** Allocation bias can result from the way that treatment and control groups are assembled. It may occur if subjects are assigned to the study groups of a clinical trial in a non-random fashion. For instance, in a study comparing oral NSAIDs

extensive diagnostic investigation and increase the probability that a disease is identified. For instance, patients who smoke may undergo increased imaging surveillance due to their smoking status, which would detect more cases of cancer in general.

**(Choice C)** Lead-time bias occurs when a screening test diagnoses a disease earlier than it would have appeared by natural history so that the time from diagnosis until death is prolonged. Affected studies may not accurately reflect an improvement in survival or alteration of the natural history.

**(Choice D)** Allocation bias can result from the way that treatment and control groups are assembled. It may occur if subjects are assigned to the study groups of a clinical trial in a non-random fashion. For instance, in a study comparing oral NSAIDs and intraarticular corticosteroid injections for the treatment of osteoarthritis, obese patients may be preferentially assigned to the corticosteroid group.

**(Choice F)** Sampling bias occurs due to a non-random sampling of a population. It can lead to a study population having characteristics that differ from the target population. A common example is that severely ill patients are the most likely to enroll in cancer trials, leading to results that are not applicable to patients with less advanced cancers.

**Educational objective:**
Recall bias results from inaccurate recall of past exposure by people in the study and applies mostly to retrospective studies such as case-control studies. People who have suffered an adverse event are more likely to recall risk factors than those without adverse experiences. Like all sources of bias, recall bias is a threat to the validity of a study.

Time Spent: 10 seconds



Last updated: [10/14/2015]

Pulmonary capillary wedge pressure (PCWP) measurements can be used to estimate left atrial pressure; the normal range is between 6-12 mm Hg, and recorded values are whole numbers. A patient in the intensive care unit has 20 serial PCWP measurements taken over the course of 2 hours. Among these 20 observations, the maximal recorded value is 12 mm Hg and the minimal recorded value is 10 mm Hg. If the next measurement is 26 mm Hg, which of the following is most likely to remain unchanged?

- A. Mean
- B. Mode
- C. Range
- D. Standard deviation
- E. Variance

Pulmonary capillary wedge pressure (PCWP) measurements can be used to estimate left atrial pressure; the normal range is between 6-12 mm Hg, and recorded values are whole numbers. A patient in the intensive care unit has 20 serial PCWP measurements taken over the course of 2 hours. Among these 20 observations, the maximal recorded value is 12 mm Hg and the minimal recorded value is 10 mm Hg. If the next measurement is 26 mm Hg, which of the following is most likely to remain unchanged?

- A. Mean [4%]
- ✔ B. Mode [89%]
- C. Range [2%]
- D. Standard deviation [4%]
- E. Variance [2%]

**Explanation:**

An **outlier** is defined as an extreme and unusual value observed in a dataset. It may be the result of a recording error, a measurement error, or a natural phenomenon. In this case, the value of 26 mm Hg is an outlier as all the other values lie between 10 and 12 mm Hg. An outlier can affect measures of central tendency (mean, median, mode) as well as measures of dispersion (standard deviation, variance).

This patient had 20 initial measurements that were all between 10 and 12 mm Hg. Although it is not required to answer this question correctly, a numerical example can help to explain the effect of an outlier such as 26 mm Hg. Assume that the patient's 20 initial measurements were: 9 measurements of 10 mm Hg, 6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
• 20
• 21
• 22
• 23
• 24
• 25
• 26

**Explanation:**

An **outlier** is defined as an extreme and unusual value observed in a dataset. It may be the result of a recording error, a measurement error, or a natural phenomenon. In this case, the value of 26 mm Hg is an outlier as all the other values lie between 10 and 12 mm Hg. An outlier can affect measures of central tendency (mean, median, mode) as well as measures of dispersion (standard deviation, variance).

This patient had 20 initial measurements that were all between 10 and 12 mm Hg. Although it is not required to answer this question correctly, a numerical example can help to explain the effect of an outlier such as 26 mm Hg. Assume that the patient's 20 initial measurements were: 9 measurements of 10 mm Hg, 6 measurements of 11 mm Hg, and 5 measurements of 12 mm Hg. The **mode** is the most frequently observed data point, so the mode for this initial dataset is 10 mm Hg (because this value is observed 9 times whereas the other 2 values are observed 6 times and 5 times). If the next measurement is 26 mm Hg, the mode would remain 10 mm Hg because this would still be the most commonly observed value. Therefore, the mode would remain unchanged. The mode tends to be resistant to outliers because outliers are not likely to affect the most frequently observed value in a dataset.

**(Choice A)** The **mean** (or average) can be quite sensitive to outliers, especially in the case of small datasets and extreme outliers. For instance, the mean of the 20 initial measurements is 10.8 mm Hg (obtained by calculating [9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 20). With an additional measurement of 26 mm Hg, the mean would become 11.5 mm Hg (obtained by calculating [1 × 26 mm Hg + 9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 21).

**(Choice C)** The **range** is equal to the maximal value minus the minimal value and

(because this value is observed 9 times whereas the other 2 values are observed 6 times and 5 times). If the next measurement is 26 mm Hg, the mode would remain 10 mm Hg because this would still be the most commonly observed value. Therefore, the mode would remain unchanged. The mode tends to be resistant to outliers because outliers are not likely to affect the most frequently observed value in a dataset.

**(Choice A)** The **mean** (or average) can be quite sensitive to outliers, especially in the case of small datasets and extreme outliers. For instance, the mean of the 20 initial measurements is 10.8 mm Hg (obtained by calculating [9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 20). With an additional measurement of 26 mm Hg, the mean would become 11.5 mm Hg (obtained by calculating [1 × 26 mm Hg + 9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 21).

**(Choice C)** The **range** is equal to the maximal value minus the minimal value and would clearly be affected by outliers. With the 20 initial measurements, the range is 12 mm Hg – 10 mm Hg = 2 mm Hg. If the next measurement is 26 mm Hg, then the range will be 26 mm Hg – 10 mm Hg = 16 mm Hg.

**(Choices D and E)** The standard deviation and variance are measures of dispersion, which reflect how spread out the values in a dataset are from one another. The standard deviation and variance tend to be sensitive to outliers because outliers increase the dispersion of datasets.

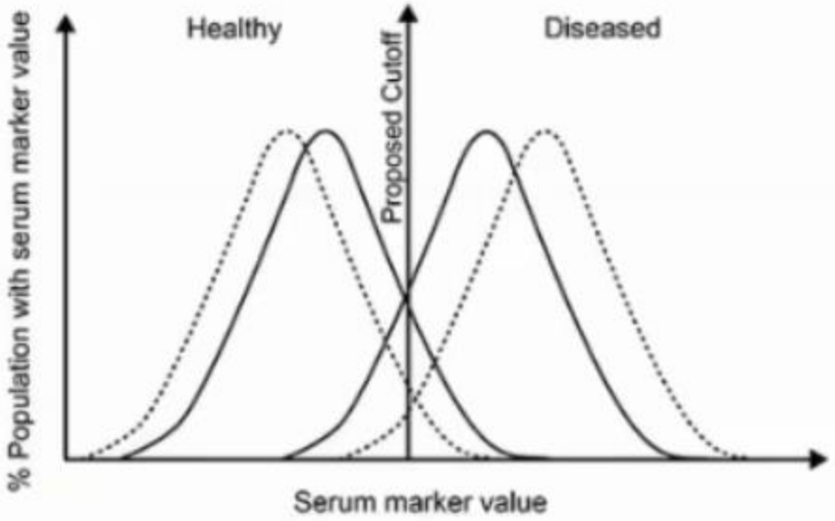
**Educational objective:**
An outlier is defined as an extreme and unusual observed value in a dataset. It can affect measures of central tendency (mean, median, mode) as well as measures of dispersion (standard deviation, variance). Modes tend to be resistant to outliers.

Time Spent: 10 seconds



Last updated: [7/29/2015]

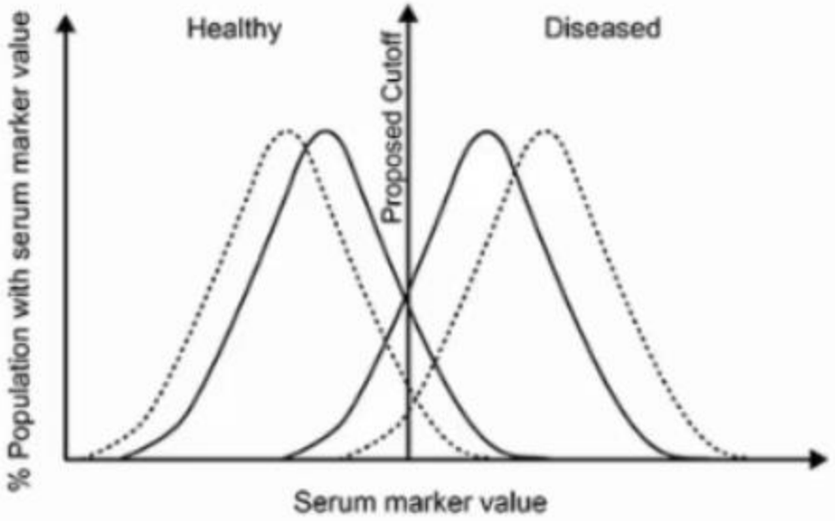
A screening test using a new serum marker is developed for diagnosing ovarian cancer. Healthy and diseased population curves along the screening dimension, along with the proposed serum marker cutoff value, are given below.




Compared to the solid curves, the dashed curves are associated with:

- A. Higher sensitivity and higher specificity
- B. Higher sensitivity and lower specificity
- C. Higher sensitivity and same specificity
- D. Lower sensitivity and higher specificity
- E. Lower sensitivity and lower specificity

A screening test using a new serum marker is developed for diagnosing ovarian cancer. Healthy and diseased population curves along the screening dimension, along with the proposed serum marker cutoff value, are given below.



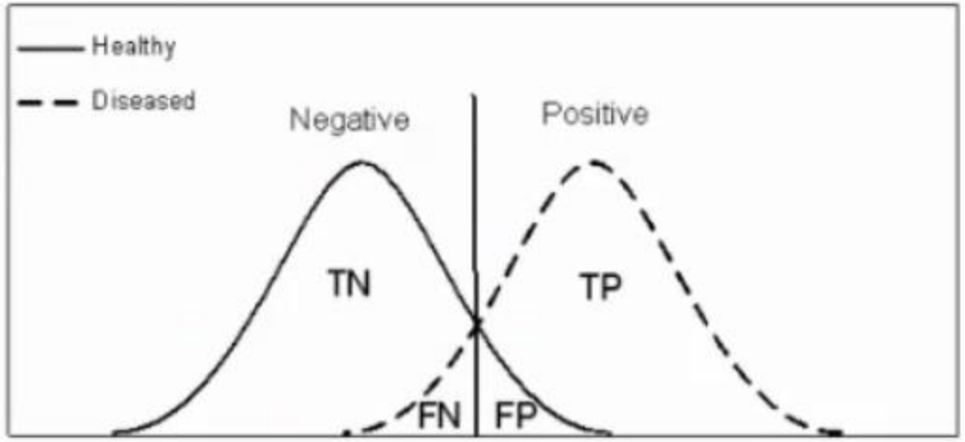

Compared to the solid curves, the dashed curves are associated with:

- ✔ (●) A. Higher sensitivity and higher specificity [62%]
- ( ) B. Higher sensitivity and lower specificity [14%]
- ( ) C. Higher sensitivity and same specificity [3%]
- ( ) D. Lower sensitivity and higher specificity [15%]
- ( ) E. Lower sensitivity and lower specificity [6%]

**Explanation:**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 • 21 • 22 • 23 • 24 • 25 • 26

**Explanation:**




FN = false negative, FP = false positive, TN = true negative, TP = true positive

The degree of overlap between the healthy and diseased population curves limits the maximum combined sensitivity and specificity of a test. In this example, the cutoff value is optimally placed to maximize both sensitivity and specificity. Sensitivity represents the ability of a test to identify those with a given disease. It is calculated as TP divided by the number of people with the disease (TP + FN). Specificity represents the ability of a test to exclude those from having a given disease. It is calculated as TN divided by the number of people without the disease (TN + FP). In this specific instance, decreased overlap between the healthy and diseased population curves decreases both the number of FP and FN, and thus allow for a test with both higher sensitivity and specificity. Therefore, the dashed curves are associated with higher sensitivity and specificity.

Depending upon the disease or condition being tested for, sensitivity or specificity may be preferred and the cutoff value adjusted accordingly. In the above example

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
• 21
• 22
• 23
• 24
• 25
• 26

FN = false negative, FP = false positive, TN = true negative, TP = true positive

The degree of overlap between the healthy and diseased population curves limits the maximum combined sensitivity and specificity of a test. In this example, the cutoff value is optimally placed to maximize both sensitivity and specificity. Sensitivity represents the ability of a test to identify those with a given disease. It is calculated as TP divided by the number of people with the disease (TP + FN). Specificity represents the ability of a test to exclude those from having a given disease. It is calculated as TN divided by the number of people without the disease (TN + FP). In this specific instance, decreased overlap between the healthy and diseased population curves decreases both the number of FP and FN, and thus allow for a test with both higher sensitivity and specificity. Therefore, the dashed curves are associated with higher sensitivity and specificity.

Depending upon the disease or condition being tested for, sensitivity or specificity may be preferred and the cutoff value adjusted accordingly. In the above example, moving the cutoff value to the right (higher value) would increase specificity at the expense of sensitivity, while moving the cutoff to the left (lower value) would increase sensitivity at the expense of specificity. A cutoff value just outside the overlapping portion would maximize the sensitivity OR specificity at 100%. However, the corresponding specificity/sensitivity of such a maximized test is dependent on the degree of overlap between the healthy and diseased population curves.

**Educational Objective:**
The degree of overlap between the healthy and diseased population curves limits the maximum combined sensitivity and specificity of a test (the area under its ROC curve). The degree to which sensitivity or specificity is affected depends upon the chosen cutoff value.

Time Spent: 6 seconds



Last updated: [8/4/2015]

A 56-year-old female comes to the emergency department because of chest pain and shortness of breath. You use test X to rule out the possibility of a pulmonary embolus. Her result is negative. Test X has a specificity of 80% and a sensitivity of 90% when tested in 100 subjects with pulmonary embolus and 100 subjects without pulmonary embolus. Assume that this patient's pretest probability for having a pulmonary embolus is equivalent to the disease prevalence in the study population. What is the probability that this patient truly does not have a pulmonary embolus?

- A. 18%
- B. 53%
- C. 66%
- D. 82%
- E. 89%

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 • 22 • 23 • 24 • 25 • 26

A 56-year-old female comes to the emergency department because of chest pain and shortness of breath. You use test X to rule out the possibility of a pulmonary embolus. Her result is negative. Test X has a specificity of 80% and a sensitivity of 90% when tested in 100 subjects with pulmonary embolus and 100 subjects without pulmonary embolus. Assume that this patient's pretest probability for having a pulmonary embolus is equivalent to the disease prevalence in the study population. What is the probability that this patient truly does not have a pulmonary embolus?

- A. 18% [6%]
- B. 53% [4%]
- C. 66% [8%]
- D. 82% [25%]
- ✔ E. 89% [57%]

**Explanation:**

This question asks for the negative predictive value (NPV) of test X. In order to determine the NPV, it helps to create a two-by-two table. The question tells us that there are 100 subjects with pulmonary embolus (PE) and 100 subjects without PE, and to assume that the patient's pre-test probability (essentially the disease prevalence modified by the physician's clinical judgment) is equal to the disease prevalence in the test population. Knowing that specificity = d/[b+d] (true negatives/total disease negatives) and sensitivity = a/[a+c] (true positives/total disease positives), we can construct the following two-by-two table:

Generic Test X

**Explanation:**

This question asks for the negative predictive value (NPV) of test X. In order to determine the NPV, it helps to create a two-by-two table. The question tells us that there are 100 subjects with pulmonary embolus (PE) and 100 subjects without PE, and to assume that the patient's pre-test probability (essentially the disease prevalence modified by the physician's clinical judgment) is equal to the disease prevalence in the test population. Knowing that specificity = d/[b+d] (true negatives/total disease negatives) and sensitivity = a/[a+c] (true positives/total disease positives), we can construct the following two-by-two table:

Generic

| | Disease positive | Disease negative | |
|---|---|---|---|
| Test positive | a<br>True positive | b<br>False positive | a+b |
| Test negative | c<br>False negative | d<br>True negative | c+d |
| | a+c | b+d | |

Test X

| | + PE | - PE | |
|---|---|---|---|
| Test X positive | 90 | 20 | 110 |
| Test X negative | 10 | 80 | 90 |
| | 100 | 100 | |

The NPV is the probability that a patient truly does not have a disease if she tests negative for the disease. Thus, NPV = d/[c+d] (true negatives/total negative tests). In this study, the NPV = 80/(10+80), or ~89%. Unlike sensitivity and specificity, NPV varies based upon disease prevalence and is inversely proportional to the prevalence of a disease. For example, NPV decreases as the prevalence of a disease increases.

**Educational objective:**

Negative predictive value (NPV) represents the probability of not having a disease

there are 100 subjects with pulmonary embolus (PE) and 100 subjects without PE, and to assume that the patient's pre-test probability (essentially the disease prevalence modified by the physician's clinical judgment) is equal to the disease prevalence in the test population. Knowing that specificity = d/[b+d] (true negatives/total disease negatives) and sensitivity = a/[a+c] (true positives/total disease positives), we can construct the following two-by-two table:

Generic

| | Disease positive | Disease negative | |
|---|---|---|---|
| Test positive | a<br>True positive | b<br>False positive | a+b |
| Test negative | c<br>False negative | d<br>True negative | c+d |
| | a+c | b+d | |

Test X

| | + PE | - PE | |
|---|---|---|---|
| Test X positive | 90 | 20 | 110 |
| Test X negative | 10 | 80 | 90 |
| | 100 | 100 | |

The NPV is the probability that a patient truly does not have a disease if she tests negative for the disease. Thus, NPV = d/[c+d] (true negatives/total negative tests). In this study, the NPV = 80/(10+80), or ~89%. Unlike sensitivity and specificity, NPV varies based upon disease prevalence and is inversely proportional to the prevalence of a disease. For example, NPV decreases as the prevalence of a disease increases.

**Educational objective:**
Negative predictive value (NPV) represents the probability of not having a disease given a negative test result. Unlike sensitivity and specificity, NPV varies based upon disease prevalence and is inversely proportional to the prevalence of a disease.

Time Spent: 9 seconds



Last updated: [9/24/2015]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
• 23
• 24
• 25
• 26

400 women aged 20-35 coming for routine check-up are asked about their smoking status. 40% of the women are smokers. Over the next ten years, 25 smokers and 24 non-smokers developed breast cancer. Which of the following best describes the study design?

- A. Prospective cohort study
- B. Retrospective cohort study
- C. Case-control study
- D. Cross-sectional study
- E. Randomized clinical trial

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
• 23
• 24
• 25
• 26

400 women aged 20-35 coming for routine check-up are asked about their smoking status. 40% of the women are smokers. Over the next ten years, 25 smokers and 24 non-smokers developed breast cancer. Which of the following best describes the study design?

- ✔ (●) A. Prospective cohort study [84%]
- ( ) B. Retrospective cohort study [4%]
- ( ) C. Case-control study [6%]
- ( ) D. Cross-sectional study [4%]
- ( ) E. Randomized clinical trial [1%]

**Explanation:**

The scenario described above is a good example of prospective cohort study. Initially a group of subjects is selected (i.e., cohort) and their exposure status is determined (smoker/non-smoker). The cohort is then followed for a certain period of time and observed for development of the outcome (breast cancer). Sometimes, the exposure status is determined retrospectively and then patients are tracked from that point of time, typically using medical records **(Choice B)**.

**(Choice C)** A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

**(Choice D)** A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and outcome. It is a

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 • 23 • 24 • 25 • 26

○ E. Randomized clinical trial [1%]

**Explanation:**

The scenario described above is a good example of prospective cohort study. Initially a group of subjects is selected (i.e., cohort) and their exposure status is determined (smoker/non-smoker). The cohort is then followed for a certain period of time and observed for development of the outcome (breast cancer). Sometimes, the exposure status is determined retrospectively and then patients are tracked from that point of time, typically using medical records **(Choice B)**.

**(Choice C)** A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

**(Choice D)** A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and outcome. It is a snapshot study design that frequently uses surveys. They are relatively inexpensive and easy to perform.

**(Choice E)** A randomized clinical trial directly compares two or more treatments. Usually, the subjects are randomly assigned to an exposure (e.g., a medication) or placebo and then followed for the development of the outcome of interest.

**Educational Objective:**
Prospective cohort studies are organized by selecting a group of individuals (i.e., cohort), determining their exposure status, and then following them over time for development of the disease of interest.

Time Spent: 7 seconds



Last updated: [9/6/2015]

A new serologic test has been developed for the detection of active pulmonary tuberculosis and is compared to the gold standard of sputum mycobacterial culture. A total of 1000 subjects are randomly selected for testing from a population with a high prevalence of tuberculosis. Results of the study are given below:

| | Sputum culture positive | Sputum culture negative | |
|---|---|---|---|
| Serologic test positive | 130 | 60 | 190 |
| Serologic test negative | 50 | 760 | 810 |
| | 180 | 820 | 1000 |

Which of the following is the positive predictive value of the test under study?

- A. 130/180
- B. 130/190
- C. 180/1000
- D. 760/810
- E. 760/820

A new serologic test has been developed for the detection of active pulmonary tuberculosis and is compared to the gold standard of sputum mycobacterial culture. A total of 1000 subjects are randomly selected for testing from a population with a high prevalence of tuberculosis. Results of the study are given below:

| | Sputum culture positive | Sputum culture negative | |
|---|---|---|---|
| Serologic test positive | 130 | 60 | 190 |
| Serologic test negative | 50 | 760 | 810 |
| | 180 | 820 | 1000 |

Which of the following is the positive predictive value of the test under study?

- A. 130/180 [11%]
- ✔ B. 130/190 [86%]
- C. 180/1000 [2%]
- D. 760/810 [0%]
- E. 760/820 [0%]

**Explanation:**

The **positive predictive value (PPV)** of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV is calculated as the proportion of subjects who truly have the disease among all those

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 • 24 • 25 • 26

**Explanation:**

The **positive predictive value (PPV)** of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV is calculated as the proportion of subjects who truly have the disease among all those with a positive test result. In contrast, the **negative predictive value (NPV)** of a test answers the question: If the test result is negative, what is the probability that a patient does not have the disease? It is calculated as the proportion of subjects who are truly free of disease among all those with a negative test result. Predictive values are of prime importance to physicians because, in clinical practice, patients will present more often with a positive or negative test than with a defined diseased or disease-free state.

Consider the following 2 × 2 (contingency) table:

| | **Positive condition** | **Negative condition** | |
|---|---|---|---|
| **Positive test result** | TP | FP | PPV=TP/(TP+FP) |
| **Negative test result** | FN | TN | NPV=TN/(TN+FN) |
| | Sensitivity = TP/(TP+FN) | Specificity = TN/(TN+FP) | |

**FN** = false negative; **FP** = false positive; **TN** = true negative; **TP** = true positive; **NPV** = negative predictive value; **PPV** = positive predictive value.

| Sensitivity = TP/(TP+FN) | Specificity = TN/(TN+FP) |
|---|---|

**FN** = false negative; **FP** = false positive; **TN** = true negative; **TP** = true positive; **NPV** = negative predictive value; **PPV** = positive predictive value.



PPV = TP / (TP + FP), where TP is true positives and FP is false positives.

NPV = TN / (TN + FN), where TN is true negatives and FN is false negatives.

In this case, PPV = 130 / (130 + 60) = 130/190.

**(Choices A and E)** Sensitivity is a test's ability to correctly identify individuals with the disease. Sensitivity = TP / (TP + FN) = 130 / (130 + 50) = 130/180. Specificity is a test's ability to correctly identify individuals without the disease. Specificity = TN / (TN + FP) = 760 / (760 + 60) = 760/820.

**(Choice C)** The total number of individuals who have tuberculosis (based on the gold standard of positive sputum culture) is 180. The total number of individuals in this sample is 1000. Therefore, the prevalence of tuberculosis in this sample is 180/1000.

**(Choice D)** NPV = 760 / (760 + 50) = 760/810.

**Educational objective:**
The positive predictive value (PPV) of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV = true positives / (true positives + false positives).

Time Spent: 10 seconds



Last updated: [8/13/2015]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
• 25
• 26

A study compared drug A versus standard therapy in preventing recurrent pulmonary embolism (PE). The absolute risk reduction for drug A versus standard therapy was 4%. The incidence of recurrent PE in the standard therapy group was 6%. There were 24 patients who developed recurrent PE in the drug A group. How many total subjects were there in the drug A group?

- A. 600
- B. 900
- C. 1200
- D. 1500
- E. 1800
- F. 2100

A study compared drug A versus standard therapy in preventing recurrent pulmonary embolism (PE). The absolute risk reduction for drug A versus standard therapy was 4%. The incidence of recurrent PE in the standard therapy group was 6%. There were 24 patients who developed recurrent PE in the drug A group. How many total subjects were there in the drug A group?

- A. 600 [17%]
- B. 900 [6%]
- ✔ C. 1200 [67%]
- D. 1500 [4%]
- E. 1800 [3%]
- F. 2100 [2%]

**Explanation:**

This question is meant to challenge your knowledge of how absolute risk reductions (ARR) are calculated. The ARR equals the event rate in the control group ($ER_{control}$) minus the event rate in the treatment group ($ER_{treatment}$). In this example, the event rate represents the incidence of recurrent pulmonary embolism (PE); $ER_{treatment}$ is the incidence of recurrent PE in the drug A group, and $ER_{control}$ is the incidence in the standard therapy group. $ER_{treatment}$ can be determined from the ARR (given as 4%) and $ER_{control}$ (given as 6%):

$$ARR = ER_{control} - ER_{treatment}$$
$$4\% = 6\% - ER_{treatment}$$

**Explanation:**

This question is meant to challenge your knowledge of how absolute risk reductions (ARR) are calculated. The ARR equals the event rate in the control group ($ER_{control}$) minus the event rate in the treatment group ($ER_{treatment}$). In this example, the event rate represents the incidence of recurrent pulmonary embolism (PE); $ER_{treatment}$ is the incidence of recurrent PE in the drug A group, and $ER_{control}$ is the incidence in the standard therapy group. $ER_{treatment}$ can be determined from the ARR (given as 4%) and $ER_{control}$ (given as 6%):

$ARR = ER_{control} - ER_{treatment}$
$4\% = 6\% - ER_{treatment}$
$ER_{treatment} = 2\% = 0.02$

This value ($ER_{treatment}$) also represents the number of events in the treatment arm divided by the number of subjects in the treatment arm. Therefore, knowing the total number of events in the treatment arm (24 instances of recurrent PE in the drug A group), the number of subjects in the treatment arm can be easily calculated:

$ER_{treatment}$ = Number of events in the treatment arm / Number of subjects in the treatment arm
0.02 = 24 / Number of subjects in the treatment arm
Number of subjects in the treatment arm = 24 / 0.02 = 1200

**Educational objective:**
Absolute risk reduction = event rate in the control group − event rate in the treatment group.

Time Spent: 9 seconds



Last updated: [8/10/2015]

An experimental drug X is being tested for the treatment of stage IV solid tumor malignancies. Part of the drug company's evaluation process is to analyze survival data after 3 months of treatment. The results are given in the table below.

| | Treated with drug X | Treated with placebo |
|---|---|---|
| Alive at 3 months | 20 | 38 |
| Dead at 3 months | 60 | 38 |

Which of the following best represents the number needed to harm for drug X?

- A. 2
- B. 3
- C. 4
- D. 7
- E. 9

An experimental drug X is being tested for the treatment of stage IV solid tumor malignancies. Part of the drug company's evaluation process is to analyze survival data after 3 months of treatment. The results are given in the table below.

| | Treated with drug X | Treated with placebo |
|---|---|---|
| Alive at 3 months | 20 | 38 |
| Dead at 3 months | 60 | 38 |

Which of the following best represents the number needed to harm for drug X?

- A. 2 [15%]
- B. 3 [25%]
- ✔ C. 4 [53%]
- D. 7 [5%]
- E. 9 [2%]

**Explanation:**

Unfortunately, not all interventions end up helping patients. The **number needed to harm (NNH)** represents the number of people who must be treated before an adverse event occurs. It is calculated in a manner similar to the number needed to treat (NNT) but using the absolute risk increase (sometimes referred to as the attributable risk) instead of the absolute risk reduction:

**Explanation:**

Unfortunately, not all interventions end up helping patients. The **number needed to harm (NNH)** represents the number of people who must be treated before an adverse event occurs. It is calculated in a manner similar to the number needed to treat (NNT) but using the absolute risk increase (sometimes referred to as the attributable risk) instead of the absolute risk reduction:

**NNH = 1 / Attributable risk**

To determine the attributable risk, first calculate the adverse event rates in the treatment and placebo groups. Note that the data is NOT presented in the standard format of a contingency (2 × 2) table, so care should be exercised in selecting the appropriate values and applying the formulas. In this case, the adverse event of interest is death at 3 months. There were 20 + 60 = 80 people treated with drug X; of those, 60 were dead at 3 months. Similarly, there were 38 + 38 = 76 people treated with placebo; of those, 38 were dead at 3 months. Therefore:

Adverse event rate in treatment group = 60 / 80 = 0.75 (ie, 75%)
Adverse event rate in placebo group = 38 / 76 = 0.50 (ie, 50%)

The attributable risk can then be calculated by subtracting the adverse event rate in the placebo group from the adverse event rate in the treatment group:

$$\text{Attributable risk} = \text{Event rate}_{\text{treatment}} - \text{Event rate}_{\text{placebo}}$$
Attributable risk = 0.75 – 0.50 = 0.25 (ie, 25%)

The absolute risk increase attributable to the treatment is 25% (75% risk of being dead at 3 months in the drug X group compared to 50% in the placebo group). The NNH then is simply the inverse of the attributable risk:

The absolute risk increase attributable to the treatment is 25% (75% risk of being dead at 3 months in the drug X group compared to 50% in the placebo group). The NNH then is simply the inverse of the attributable risk:

NNH = 1 / Attributable risk
NNH = 1 / 0.25 = 4

This result indicates that for every 4 patients treated with drug X, 1 will experience an adverse event (in this case, death).

An alternative solution involves reordering the data into a standard contingency table format, with the exposure of interest (drug X) in the topmost row and the outcome of interest (dead at 3 months) in the leftmost column:

| | Dead at 3 months | Alive at 3 months |
|---|---|---|
| Treated with drug X | a = 60 | b = 20 |
| Treated with placebo | c = 38 | d = 38 |

The formula for attributable risk is: [ a / (a + b) ] – [ c / (c + d) ] = [ 60 / (60 + 20) ] – [ 38 / (38 + 38) ] = 60/80 – 38/76 = 0.75 – 0.50 = 0.25; so NNH = 1 / 0.25 = 4.

**Educational objective:**
Number needed to harm = 1 / Attributable risk.

Time Spent: 9 seconds



Last updated: [7/30/2015]